Structural Estimation of Labor Adjustment Costs with Temporally Disaggregated Data and Gross Employment Flows

Fırat Yaman*

Abstract

Estimating labor adjustment costs is plagued by a variety of errors, many arising from data limitations. Most researchers have assumed that adjustment decisions are made at the firm level, that adjustment happens at the frequency at which a firm is observed (typically annually or quarterly), and that adjustment costs are incurred on net changes in employment. In this paper, I estimate a dynamic optimization model of labor adjustment of establishments based on data that permit 1) specifying any desired adjustment frequency, 2) estimating the model based on net and on gross employment flows and 3) allowing for simultaneous hirings and separations. The unit of observation is an establishment. Results for adjustment costs depend crucially on the model specification. Only a monthly adjustment model yields cost parameters in a reasonable range, while estimates from quarterly and annual adjustment models imply negative (adjustment implies a gain rather than a loss) or excessive adjustment costs. Estimating the model on net employment changes implies hiring and separation costs of around four annual median salaries, while the model on gross changes implies costs on the order of 1.7 annual median salaries. Adjustment costs differ significantly between small and large establishments. However, a dynamic model performs only marginally better than a static model with respect to out-of-sample predictions.

JEL classification: C25; D22; J23

Keywords: Dynamic discrete choice; Adjustment costs; Labor demand; Temporal aggregation

*Department of Economics, City University London. Email: Firat.Yaman.1@city.ac.uk.

1 Introduction

Estimating the magnitude and structure of labor adjustment costs has a long tradition in economics and has faced a number of challenges.[1] Two main approaches can be distinguished. The more straightforward approach is to use firm and establishment surveys on the costs of vacancy posting, training, redundancy payments etc. and relating those costs to the number of hires and separations (as in Abowd and Kramarz (2003) and Kramarz and Michaud (2010)). This information might not always be available, or the costs reported by firms might not fully account for the loss of profits due to disruptions in the production process. For those reasons most contributions to this literature analyze models of profit-maximizing firms and conduct a structural estimation where the cost parameters are treated as estimable. The parameter values are chosen (estimated) to reproduce certain observed data moments and statistics or employment trajectories of the firms as closely as possible.

The structural approach suffers from a number of misspecification problems. In this paper I focus on two of them. One is the question of the frequency at which firms revise their employment decisions. Here, data availability dictates what the researcher assumes. Since employment data has usually come from firm and establishment surveys, researchers have only known stocks of employment at fixed time intervals (e.g. at the start of every quarter), and had to assume that firms make hiring and separation decisions at the same frequency as the data was available. But the choice of adjustment frequency is not innocuous for the estimation. A firm is more likely to change its employment in the long run. Thus, an infrequently observed firm is likely to exhibit no or few inactive periods, suggesting that adjustment costs are (marginally) small. Using higher-frequency data (quarters), Varejão and Portugal (2007) conclude that "micro and quarterly (or more frequent) data are essential for studying the dynamics of factor adjustment because aggregation (spatial or temporal) smoothes away any nonlinearities present at the plant or firm levels." Bloom (2009) also reports a high sensitivity of adjustment cost estimates to adjustment frequency. The issue of temporal aggregation is also identified in Addison et al. (2014) as a major outstanding issue in labor demand research.

A more conceptual problem with the structural approach is to estimate costs based on net worker flows. This necessarily follows from assuming that labor is homogenous. If all workers are the same, then why would a firm decide to hire and fire simultaneously? Yet, the phenomenon of churning - the turnover of

[1] I examine labor adjustment, but the problems associated with estimating capital (or any factor) adjustment costs are almost identical. Hamermesh and Pfann (1996) is an excellent survey for the literature until 1996 and addresses all of the points discussed in this paper.

workers over and above job creations or destructions - is pervasive, even within narrowly defined classes of employees (Abowd et al. (1999) and Burgess et al. (2001)). Structural models accommodate churning through worker-initiated separations, but that still underestimates - perhaps substantially - employment adjustments initiated by the firm. Here, too, the modelling choices matter for the estimation. By using only net flows, the researcher is measuring the main choice variable of the model with error.

In this paper I address the problem of temporal aggregation by using continuous employment data, including all in- and outflows of employees, from a linked employer-employee dataset. I estimate my model at different adjustment frequencies and compare the resulting cost estimates. I also use gross instead of net adjustments and rationalize this choice by developing a model of dynamic discrete choice which accommodates the simultaneous hiring of and separating from workers as an establishment's choice. To my knowledge, there is no other model of adjustment costs that does this. I show that these choices - temporal frequency, net or gross adjustment - have substantial effects on the resulting cost estimates. The limitation of this paper is that I do not offer a set of "correct" cost estimates. Developing a convincing test for discriminating between the myriad of cost estimates that we have based on modelling choices remains an open and very challenging question.

I build a model of linear hiring and separation costs which accommodates churning, the observation of simultaneous hirings and separations, and estimate it for different aggregation levels, and separately for adjustment costs defined over gross versus net flows. I show that the most important factor contributing to higher adjustment cost estimates are the number of observations of inactivity. The economically most sensible estimates of adjustment costs are obtained for a model of 1) monthly adjustment on 2) gross employee flows and figure around 35,000 Euros per hire and per separation. Without an external source on the magnitude of adjustment costs there is nothing to compare these estimates against.[2]

Despite the above mentioned difficulties economists should be interested in estimating adjustment costs. Macroeconomists will be interested in the effect of adjustment costs on unemployment and the labor market effects of the business cycle; see, for example Bentolila and Bertola (1990) and Hopenhayn and Rogerson (1993). Policymakers will be interested in the role of adjustment costs when considering labor market interventions such as wage subsidies in a recession (such as the German *Kurzarbeit* program), and economists

[2]Unfortunately, but not unsurprisingly, estimates in the literature have a very wide range. For Germany, Mühlemann and Pfeifer (2013) report hiring costs between 4,000 and 6,000 Euros per hire. However, Hamermesh and Pfann (1996) cite some accounting studies which find adjustment costs of one year of payroll costs for the average worker. More on this in the discussion section.

working in industrial organization should know the right specification of a firm's objective function for purposes such as estimating a production function. Labor economists are, I presume, intrinsically interested.

2 Literature

The literature on labor adjustment costs can be categorized on a number of dimensions. The first is the choice of the unit that makes hiring and separation decisions ("spatial" aggregation). Hamermesh (1989) shows that at the plant level long spells of inactivity are followed by lumpy labor adjustment – indicating either the presence of fixed costs of adjustment or non-differentiability at a net adjustment of zero; see, for example, Abel and Eberly (1994). Aggregated to the firm level, labor adjustment appears to be much more frequent and smooth, suggesting a convex adjustment cost function. Researchers who have been interested in the adjustment costs per se, rather than using them as a device for a better model fit, have employed plant or firm-level data over higher aggregates ever since.

Second, a similar problem occurs with "temporal" aggregation. A plant or firm is more likely to change its employment in the long run. Thus, an infrequently observed plant is likely to exhibit no or few inactive periods with the same implications for cost estimates as in the spatial aggregation case, as demonstrated by Varejão and Portugal (2007) who use quarterly data. Bloom (2009) highlights the same point. His model is simulated on monthly adjustment frequency and aggregated to annual measures to match them with annual data. He shows that cost estimates are lower if annual adjustment frequencies are assumed. It is interesting to note that this point is often not even discussed in the literature and that there is not a consensus. For example, using annual industry-level data, Hall (2004) argues that "both time aggregation and aggregation across firms is probably not an important source of bias in estimation". He finds no evidence for labor adjustment costs: For 10 out of 18 industries the adjustment cost parameter carries the "wrong" sign.

Third is the correct specification of the cost function. Hamermesh and Pfann (1996) reject the hypothesis of convex adjustment costs in favor of other types of costs such as fixed costs, linear costs, and production disruption costs.[3]

Forth is whether to use net or gross employment flows. When only the stock of employment at fixed

[3] One or several of those components are used by Abel and Eberly (1994); Abowd and Kramarz (2003); Aguirregabiria and Alonso-Borrego (2014); Asphjell et al. (2014); Bentolila and Bertola (1990); Bloom (2009); Cooper et al. (2015); Ejarque and Portugal (2007); Lapatinas (2009); Pfann and Palm (1993); Nilsen et al. (2007); Rota (2004); and Varejão and Portugal (2007).

intervals is reported, then net flows must be used by necessity. If adjustment costs are costs of hiring and separations rather than of job creation and destruction, then gross flows should be used as in Abowd and Kramarz (2003) and Kramarz and Michaud (2010). Abowd et al. (1999) and Burgess et al. (2001) show that net and gross flows differ considerably within an establishment, even within a skill group.

Fifth is labor heterogeneity which is closely related to the previous point. All structural estimations of adjustment costs specify a production function with one type of labor without acknowledging that in such a framework it makes no sense for an establishment to hire and fire within a period. Thus, adjustment costs in such a model can only be incurred over net changes. Goolsbee and Gross (2000) illustrate how this heterogeneity aggregation impacts on estimates of capital adjustment costs.

Sixth, a choice must be made concerning whether adjustment costs should be estimated using reported costs by firms and establishments or using a structural model. If the objective is to quantify adjustment costs, then the former is preferable and is done by Abowd and Kramarz (2003) as well as Kramarz and Michaud (2010). Still, not all costs might be captured by establishment reports. In that case, or if interest lies in comparing models with and without adjustment costs, or in counter-factual policy evaluations, structural estimation is the method of choice, which in turn necessitates another set of modelling choices. Should one use a partial equilibrium or a general equilibrium model? What is the production function? What process determines wages? What data moments should be matched? And what should be the test to discard one type of model in favor of another?

Table 1 gives a summary of the more recent literature on labor adjustment costs. The table is not exhaustive, and I chose the list of studies with an eye to the wide variety of modelling choices that characterize the literature. Comparing results across studies is difficult because the adjustment cost function and other modelling choices vary widely across the studies, but the results testify to the wide range of estimates that have been found. Some of those studies focused only on the demand side of the labor market, whereas others employ a general equilibrium (search and matching) model. Abowd and Kramarz (2003) is the only study based on reported costs, and there is a wide spectrum of estimation methods, though most use a MSM estimator. Interestingly, these are the same studies which find very low - often economically insignificant - estimates of labor adjustment costs.

Here I want to stress that apart from this paper there are two studies which use gross worker flows,

Table 1: Some recent estimates of labor adjustment costs

Paper	Country	Frequency	Gross or net	Equilibrium	Method	Other features	Estimate in % of average annual wage
Abowd & Kramarz (2003)	France	n.a.	Gross	n.a.	Regression	Based on reported costs	0.53 for terminations 0.02 for hiring
Aguirregabiria & Alonso-Borrego (2014)	Spain	Annual	Net	Partial	Discrete choice		0.15 for hiring 0.5 for firing
Asphjell et al. (2014)	Norway	Annual	Net	Partial	MSM	Includes capital adjustment	0.001 (hiring and firing)
Bloom (2009)	USA	Monthly	Net	Partial	MSM	Includes capital adjustment	9.5 (hiring and firing)
Cooper et al. (2007)	USA	Monthly	Net	General	MSM	Includes hours adjustment	0.002 for hiring 0.002 for firing
Hall (2004)	USA	Annual	Net	Partial	GMM (Euler equation)	Industry level	0 (hiring and firing)
Lapatinas (2009)	Greece	Annual	Net	Partial	MSM		0.30 (hiring and firing)
Nilsen et al (2007)	Norway	Annual	Net	Partial	Ordered Probit and Euler equation		not identified
Rota (2004)	Italy	Annual	Net	Partial	GMM (Euler equation)		1.25 fixed (hiring and firing)
Trapeznikova (2017)	Denmark	Monthly	Gross	General	MSM	Includes hours adjustment	0.12 for hiring 0 for firing (assumed)
Yaman (2018)	Germany	Monthly	Gross	Partial	Discrete Choice	Four choice regimes	1.7 for hiring 1.8 for separation
Yaman (2018)	Germany	Monthly	Net	Partial	MSM		0.08 for hiring -0.003 for separation

Note: Where applicable, I assume an establishment size of 50 and an adjustment of 5 workers to quantify adjustment costs in the last column.

of which one is not a structural estimation (Abowd and Kramarz (2003)), and the other assumes no firing costs and is in continuous time such that hirings along with firings can in principal occur over a given time period (Trapeznikova (2017)). In contrast, simultaneous hiring and firing can occur in my model because an establishment might wish to replace unproductive workers. My model is to my knowledge the only one which develops a heterogeneous worker framework to rationalize simultaneous hiring and separation decisions.

With respect to adjustment frequency, Bloom (2009) and Cooper et al. (2007) use a model with monthly adjustment, but their data require them to aggregate the simulated choices either temporally and/or across establishments to match aggregate data moments. I in contrast use *observed* monthly adjustments at the establishment level to estimate labor adjustment costs.

3 Institutional Background

The OECD (1999) characterizes Germany's labor market in the 1990s as one of the more regulated labor markets of the OECD countries. Major reforms to reduce unemployment were initiated in the first half of the 2000s, but these are not relevant to my sample period. The OECD (1999) lists the following institutional and/or legislative features related to employment protection for Germany:

1. If a firm wishes to dismiss an employee, written notice has to be given to the employee and - where applicable - to the workers' council. If an employee is dismissed against the objection of the workers' council, the case escalates to a labor court and the dismissal does not become effective until a decision is reached.
2. Required notice periods can range from 2 weeks to 7 months, depending on the employee's tenure. There is no legal severance pay entitlement, but these can be stipulated in collective agreements. For a worker with 20 years tenure, the severance pay for unfair dismissal is 18 months of wages.
3. For a dismissal to be classified as "fair", a firm must make sure that an employee cannot be retained in another capacity within the same establishment or enterprise, and "social" considerations must be taken into account (e.g. age and number of dependants).

There are two more distinct features of the German labor market which contribute to labor adjustment costs. First is the strong involvement of establishments in the training of young workers. Harhoff and Kane (1997) estimate training costs to be between $5,000 and $10,000 in 1990 values. Second is the existence of workers' councils at many establishments which need to be informed and consulted for certain decisions, including re-structuring of employment and recruitment, and enjoy certain co-determination rights for procedures related to dismissals as describe above (see also Addison et al. (2001)).

Another important difference between the German and the American labor market is that the latter is much more dynamic. Bachmann (2005) report a job-to-job transition rate for the period 1980 to 2000 of 0.82% in Germany, compared to around 2.5% in the USA (Nagypál (2008) and Fallick and Fleischman (2004)). However, movements into and out of unemployment are equally much lower in Germany than in the USA. Jung and Kuhn (2014) report employment to unemployment flows of 0.5% (2.0%) for Germany (USA) and employment to employment flows of 0.9% (2.6%).

4 Model

The main challenge in setting up an estimable model which allows simultaneous hiring and establishment-initiated separations is to allow for worker heterogeneity such that only observable characteristics of the workforce (e.g. the number of employees) enter the model as variables, while the unobservable parts (e.g. the probability distribution of productivity among employees) enter as "errors": factors which are considered by the establishment, but not observable to the researcher. Vectors (and sets) are denoted in bold characters.

Establishments are assumed infinitely-lived, maximizing the present value of current and all expected future profits, and being price takers in product and factor markets. In the German context the exogeneity of wages is partly defensible due to the large coverage of workers by collective agreements, mainly at the industry level. In 2000, 63% of workers in West and 45% of workers in East Germany fell under the coverage of a collective agreement (Kohaut and Schnabel (2003)), even though the agreed wages are only binding as a wage floor. Jung and Schnabel (2011) find that in 2006, 40% of German establishments pay wages above the agreed level. An indirect piece of evidence for wage rigidities in Germany is Glitz (2012) who finds effects of immigration on native unemployment, but not on wages.

Workers are heterogeneous in their productivity. I use the term productivity in a broad sense. Any factor which increases the marginal revenue of labor (e.g. a price increase in the product market) is an increase in productivity. The vector of worker productivities at the end of period $t-1$ is denoted $\hat{\mathbf{a}}_{t-1}$. At the start of period t, before any hiring or separation decisions are made, each worker receives a productivity shock ε_t^a, some workers might quit ($\boldsymbol{\eta}_t^q$), while workers who were temporarily absent might return to the establishment ($\boldsymbol{\eta}_t^r$, henceforth *recalls*). The productivity vector just before employment decisions are made is thus given by:[4]

$$\mathbf{a}_t = \left((\hat{\mathbf{a}}_{t-1} + \boldsymbol{\varepsilon}_t^a) \setminus \boldsymbol{\eta}_t^q\right) \cup \boldsymbol{\eta}_t^r \tag{1}$$

The number of workers changes according to

$$\ell_t = \hat{\ell}_{t-1} + \eta_t \tag{2}$$

where η is the number of recalls minus the number of quits. After observing $\mathbf{a}_t$ the establishment chooses which workers to separate from, and how many workers to hire. I assume that the establishment observes the quality of its existing workers but it cannot observe the quality of a worker it hires. However, it has an expectation over the productivity of a hire which can differ across firms and time. If the firm decides to separate from workers, then it will do so from the ones with the smallest marginal productivities. The state vector of the firm thus consists of the productivity vector $\mathbf{a}_t$, the expected productivity of new hires $E(a_t^h)$, and the wage w_t. The productivity vector after the choices are made is given by

$$\hat{\mathbf{a}}_t = (\mathbf{a}_t \setminus \mathbf{f}_t) \cup \mathbf{h}_t$$

where **f** is the set of productivities of the workers the establishment separates from, and **h** are the productivities of the new hires. The resulting level of employment is:

$$\hat{\ell}_t = \ell_t - f_t + h_t$$

[4]The notation is abusive. I am combining a vector addition $+$ with set operations $\setminus$ and $\cup$.

Sales are a (non-injective) function of the productivity vector: $S_t = f_s(\hat{\mathbf{a}}_t)$. The vectors $\mathbf{a}_t$ and $\hat{\mathbf{a}}_t$ can be fully characterized by their lengths, ℓ_t, and $\hat{\ell}_t$, and the probability distribution of the productivities, which I denote $\mu_t(a)$ and $\mu_t(\hat{a})$. The typical model found in the literature can be regarded as a special case, with $\mu_t(a)$ being a degenerate distribution, and sales being simply a function of number of employees, and a scalar productivity measure. We can characterize the state vector before employment decisions are made by

$$\mathbf{x}_t = (\mu_t(a), \ell_t, E(a_t^h), w_t)$$

The establishment observes $\mathbf{a}_t$ and $\hat{\mathbf{a}}_t$, but the researcher only observes ℓ_t and $\hat{\ell}_t$. Through equation 1, I can further characterize $\mu_t(a)$ as a vector function of $\mu_{t-1}(\hat{a})$, $\hat{\ell}_{t-1}$, $\boldsymbol{\varepsilon}_t^a$, $\boldsymbol{\eta}_t^q$, and $\boldsymbol{\eta}_t^r$. And since

$$\begin{aligned} S_{t-1} =& f_s(\hat{\mathbf{a}}_{t-1}) \\ =& f_s(\mu_{t-1}(\hat{a}), \hat{\ell}_{t-1}) \end{aligned}$$

I can write $\mu_{t-1}(\hat{a})$ as a vector function of S_{t-1}, $\hat{\ell}_{t-1}$, and an unobservable component $\boldsymbol{\varepsilon}_{t-1}^{\mu}$. The latter has to be included since, given $\hat{\ell}_{t-1}$, different productivity distributions $\mu_{t-1}(\hat{a})$ can result in the same value of sales. While knowledge of S_{t-1} and $\hat{\ell}_{t-1}$ does not uniquely identify $\mu_{t-1}(\hat{a})$, knowledge of S_{t-1}, $\hat{\ell}_{t-1}$, and $\boldsymbol{\varepsilon}_{t-1}^{\mu}$ does. Finally, I also have to consider that $\boldsymbol{\eta}_t^q$ and $\boldsymbol{\eta}_t^r$ are not observed, but the net change of employment, η_t, is, and I denote the missing information that maps η_t to $\boldsymbol{\eta}_t^q$ and $\boldsymbol{\eta}_t^r$ by $\boldsymbol{\varepsilon}_t^{\eta}$.

I can now characterize the state vector in terms of the observable variables S_{t-1}, $\hat{\ell}_{t-1}$, η, w_t, and the unobservable variables $E(a_t^h), \boldsymbol{\varepsilon}_t^a$, $\boldsymbol{\varepsilon}_{t-1}^{\mu}$, and $\boldsymbol{\varepsilon}_t^{\eta}$. I further use equation 2 to reduce the observables to $\mathbf{x}_t^o = (S_{t-1}, \ell_t, w_t)$, and collect all unobserved components in $\mathbf{x}_t^u = (E(a_t^h), \boldsymbol{\varepsilon}_t^a, \boldsymbol{\varepsilon}_{t-1}^{\mu}, \boldsymbol{\varepsilon}_t^{\eta})$. The entire state vector is denoted by $\mathbf{x}_t = (\mathbf{x}_t^o, \mathbf{x}_t^u)$.

4.1 Example

Before proceeding to the dynamic optimization problem it will be helpful to collect ideas by ways of a simple example. An establishment with two employees with productivities a_1 and a_2 has to maximize expected current profits. The wage is $\sqrt{5/8}$, both hiring and separations cost 0.1, and the expected productivity $E(a)$

Table 2: Expected profits for different scenarios and choices

Scenario	(1)	(2)	(3)	(4)
	A = 1.2	A = 1.2	A = 0.8	A = 1.5
	a = {4,6}	a = {3,7}	a = {4,6}	a = {4,6}
Inactive	**2.21**	2.21	0.95	3.16
Hiring	2.18	2.18	0.63	**3.34**
Separation	2.05	2.28	**1.07**	2.78
Hiring and separation	2.20	**2.38**	0.87	3.19

Note: Expected profits under four different scenarios. Hiring refers to hiring one worker, separation refers to separating from the least productive worker. The wage is set to $\sqrt{5/8}$. $E(a)$ of a new hire is set to 5. The expected profit is given by equation 3. The best choice for each scenario is highlighted in bold and yellow.

of a new hire is 5. The profit function is given by

$$\pi = A \times \sqrt{\sum_{i=1}^{N} a_i} - wN - 0.1f - 0.1h \tag{3}$$

where N is the number of workers. Table 2 shows profits of each choice (hiring one worker, separating from one worker, both, or none) under different scenarios for A, a_1, and a_2. The optimal decision is highlighted.

The table shows how each of the four discrete choices (being inactive, hiring one worker, separating from one worker, and replacing the least productive worker with a new hire) can be optimal under different scenarios. When establishment productivity is high (low) under scenario 4 (3), the establishment expands (contracts). A simultaneous hiring and separation occurs when one worker's productivity drops below a certain level to make it worthwhile to replace him, as in scenario 2. Note that the revenue of the establishment under scenarios 1 and 2 before any adjustment decisions are made are the same. Only the spread of worker productivities under scenario 2 is wider. In the model this difference would be absorbed in the term $\boldsymbol{\varepsilon}^{\mu}$, which are characteristics of the distribution of worker productivities which are not observed by the researcher.

4.2 Dynamic discrete choice model

The establishment's choice problem is the following:

$$V(\mathbf{x}_t) = \max_{h_t \geq 0, f_t \geq 0} \pi^e(\mathbf{x}_t, h_t, f_t) - C(h_t, f_t) + \beta \mathbb{E}_{\mathbf{x}_{t+1}|\mathbf{x}_t, h_t, f_t} V(\mathbf{x}_{t+1}) \tag{4}$$

where π^e is the expected profit function, and h_t and f_t are the (non-negative) number of hires and separations chosen by the firm. The next period is discounted at rate β, and C is the adjustment cost function. Expectations are taken over the future state conditional on the current state and decisions. The current profits are expected, as the firm only has an expectation over the quality of its new hires.

If the profit function and adjustment cost function were differentiable everywhere, then it would be easy to derive an Euler equation and to estimate the adjustment cost parameters with generalized method of moments. Evidence against an everywhere differentiable cost function is quite strong; see, for example, Hamermesh and Pfann (1996). Employment changes are relatively rare, and tend to be lumpy when they do occur, suggesting the presence of either fixed costs or non-differentiability at the point of no adjustment. Moreover, at least for small establishments, assuming a continuously adjustable level of employment is also problematic, in particular if the empirical model is set up in a way that non-activity translates into higher adjustment cost estimates, thus confounding adjustment costs with indivisibilities of labor.[5] I model the adjustment cost function to be linear in hires and separations, thus introducing a non-differentiability at zero adjustment and consequently a range for the state space that will make inactivity the optimal choice of an establishment as in Abel and Eberly (1994). An Euler equation approach with non-differentiability is still possible as demonstrated in Aguirregabiria (1997) and Cooper et al. (2010), though the requirements for consistent estimation of this Euler equation are unlikely to be met. The main problem is the endogeneity of the moment of adjustment; see Aguirregabiria (1997) for a detailed exposition.

Following the steps in Aguirregabiria (1999), the decision space of the firm can be split up into four discrete areas: Let H be the discrete choice of $h > 0$ and $f = 0$, F the choice of $h = 0$ and $f > 0$, P (for put) the choice of $h = 0$ and $f = 0$, and B (for both) the choice of $h > 0$ and $f > 0$. Denote the choice set of exhaustive and mutually exclusive choices by $D = \{H,F,P,B\}$. Let d be the discrete choice of an establishment: $d \in D$. With each of these choices, there is an associated optimal level of hires and separations. For example, for $d = H$, we would constrain $f = 0$, but h would be the solution to the problem

$$V^H(\mathbf{x_t}) = \max_{h_t>0} \pi^e(\mathbf{x_t}, h_t, f_t = 0) - C(h_t, f_t = 0) + \beta \mathbb{E}_{\mathbf{x_{t+1}}|\mathbf{x_t},h_t,f_t=0} V(\mathbf{x_{t+1}})$$

Let $\delta^d(\mathbf{x_t}) = \left[h^d(\mathbf{x_t}), f^d(\mathbf{x_t})\right]$ denote the solutions for h and f in the discrete regime d. For a current choice

[5] This point is also made in Lapatinas (2009) who models the choice of employees as a discrete choice.

d, next period's value function itself can be split into the four discrete choices so that

$$\mathbb{E}_{\mathbf{x_{t+1}}|\mathbf{x_t},d_t} V(\mathbf{x_{t+1}}) = \mathbb{E}\left[\max_{j \in D}\left(V^j(\mathbf{x_{t+1}})\right)|\mathbf{x_t}, d_t\right]$$

The "discrete" version of the establishment's problem in equation 4 is thus given by

$$V(\mathbf{x_t}) = \max_{d_t \in D} \pi^e(\mathbf{x_t}, \delta^d(\mathbf{x_t})) - C(\delta^d(\mathbf{x_t})) + \beta\mathbb{E}\left[\max_{j \in D}\left(V^j(\mathbf{x_{t+1}})\right)|\mathbf{x_t}, d_t\right]$$

and the solution to this problem is denoted $d_t^*(\mathbf{x_t})$.

To summarize, the establishment considers a discrete action (and the associated optimal h and f). In doing so, it forms expectations of the future state given its current state and this discrete action. It knows that in the next period it will again choose among the best of the four discrete options (and the associated optimal h and f).

4.3 Empirical model

Following again Aguirregabiria (1999) I decompose each component which involves $\mathbf{x}$ as the sum of an empirical counterpart which can be constructed from observable state variables and a term capturing the deviation of the individual establishment's expectation from this empirical counterpart. I model

$$\pi^e\left[\mathbf{x}_t, \delta^d(\mathbf{x}_t)\right] = \mathbb{E}\left[\pi^d(\mathbf{x}_t^o)\right] + u_t^{d,\pi} \tag{5}$$

where the first term is the expected profit under discrete choice d among establishments with observed state variables $\mathbf{x}_t^o$, and $u_t^{d,\pi}$ captures the deviation of the individual establishment's expectation from the empirical average. If the average profit expectations are equal to average profits, then $\mathbb{E}\left(u_t^{d,\pi}|\mathbf{x}_t^o\right) = 0$, which is what I assume. On average, establishments cannot be too optimistic or too pessimistic (though of course any single establishment can). Equation 5 describes the transition from theory to data, and it is important to stress that the left-hand side are expected profits from the perspective of the establishment, but the expectation operator on the right hand side refers to the empirical expectation of profits of firms with the same characteristics. Accordingly, the unobserved components of the establishment's state, $\mathbf{x}_t^u$, are conceptually different from the empirical error $u_t^{d,\pi}$. The ensuing estimation does not estimate any technological parameters relating to the profit function. But I am not interested in those parameters. This simplification comes at a cost. Some assumptions have to be made regarding $u_t^{d,\pi}$ (see below), and those assumptions might or might not be defensible. More specifically, the assumptions would hold for some profit functions, and not for others.

I decompose adjustment costs as

$$\begin{aligned} C\left[\delta^d(\mathbf{x}_t)\right] =&\mathbb{E}\left[C^d(\mathbf{x}_t^o)\right] + u_t^{d,C} \\ =&\tau^+\mathbb{E}\left(h_t|\mathbf{x}_t^o, d_t\right) + \tau^-\mathbb{E}\left(f_t|\mathbf{x}_t^o, d_t\right) + u_t^{d,C} \end{aligned}$$

so the adjustment costs are based on the average adjustments under the choice d and an unobservable component $u_t^{d,C}$. The current period profits net of adjustment costs for discrete choice d are thus

$$\begin{aligned} \pi^e\left[\mathbf{x}_t, \delta^d(\mathbf{x}_t)\right] - C\left[\delta^d(\mathbf{x}_t)\right] =&\mathbb{E}\left[S(\mathbf{x}_t^o)\right] - w_t\mathbb{E}(\hat{\ell}_t) - \tau^+\mathbb{E}(h_t|\mathbf{x}_t^o, d_t) - \tau^-\mathbb{E}(f_t|\mathbf{x}_t^o, d_t) + u_t^{d,\pi} + u_t^{d,C} \\ =&\mathbf{z}(\mathbf{x}_t^o, d_t)\theta_z' + u_t^d \end{aligned}$$

where $\mathbf{z}(\mathbf{x}_t^o, d_t) = \left(\mathbb{E}\left[S(\mathbf{x}_t^o)\right] \quad w_t\mathbb{E}(\hat{\ell}_t) \quad \mathbb{E}(h_t|\mathbf{x}_t^o, d_t) \quad \mathbb{E}(f_t|\mathbf{x}_t^o, d_t)\right)$, $\quad \theta_z = \left(1 \quad -1 \quad -\tau^+ \quad -\tau^-\right)$ and $u_t^d = u_t^{d,\pi} + u_t^{d,C}$.

With $\mathbf{x}_t^o$ containing three components, and having one u_t^d for each of the four discrete choices, this is a formidable problem to solve. Rust (1987) shows that under some assumptions the dimensionality of this problem can be greatly reduced by eliminating the u_t^d from the state space. The following standard assumptions need to be made:

1. The shocks u_t^d are independent across alternatives, across time, and follow an extreme value type I distribution.

2. The distribution of the future observable state variables is independent of u_t^d conditional on the current state variables and the discrete choice, that is $F(\mathbf{x}_{t+1}^o|d_t, \mathbf{x}_t^o, \mathbf{u_t}) = F(\mathbf{x}_{t+1}^o|d_t, \mathbf{x}_t^o)$.

In the appendix I show how this formulation results in a probabilistic formulation of the four discrete choice options, and four estimable parameters: σ_s, a metric of the measurement error in establishment sales; σ, the scale parameter of the extreme value distribution; τ^+ and τ^-, the marginal costs of hiring and separations. The estimation is carried out by maximum likelihood and is discussed below.

4.4 Discussion

The model developed in the previous section has important advantages over the class of models used in the literature. Most importantly, it can account for churning as an active choice of an establishment. Churning can also occur in models with (random) worker-initiated separations as in Bachmann et al. (2017), but in the context of adjustment costs the exact reason for churning is important if the cost of worker-initiated separations is different from the cost of establishment-initiated separations.

Second, the moments approach based on aggregate data moments suffers from a problem akin to what Berry et al. (1995) call an "embarrassing 'over fitting' problem on aggregate data". Suppose we have two establishments with the same past and current revenue, the same beginning-of-period employment, and facing the same wage. In all structural models that I am aware of the labor adjustments of those two establishments MUST have been the same, since the revenues and employment determine fully the technology variable of the model (e.g. TFP), even if the technology variable is establishment-specific. Clearly, this imposes a large degree of - counterfactual - establishment homogeneity. In my sample, looking at observations which fall into the central six percentiles on each dimension of $\mathbf{x}_t^o$, 50 out of 80 observations are inactive, but the remaining observations show hiring of up to four and separations of up to seven workers, and all remaining discrete choices (hiring, separating, and both) can be observed. The dynamic discrete choice model allows me in a simple way to account for different choices of establishments which are otherwise the same on all observable dimensions. Unlike the alternative model discussed below, the establishment's choice in the dynamic discrete choice model is not determined by the observables.

Third, the model is more parsimonious than other models used in the literature. Since the interest is in estimating labor adjustment costs, the model does not specify a production function, and has only a handful of parameters that need to be estimated.

Another advantage is that the model lends itself to estimation based on every establishment's adjustment choices rather than picking a number of aggregate statistics to match. Much more information enters the estimation than in a simulated method of moments approach. It also "relieves" the researcher from having to choose which particular moments to match.

The costs of the dynamic discrete choice model are mirror images of its advantages. The two assumptions on the error term are crucial, and especially the independence over time is unlikely to hold. Many aspects of the labor market (the production function, evolution of technological parameters, wage determination) are unspecified and result in a loss of theoretical holism and appeal. We remain agnostic about what would happen should some of those technological parameters change. Those are short-comings, and I acknowledge them as such. But there are some mitigating factors. The independence of the error terms is conditional on the observable state variables, and since the evolution of those state variables conditional on the choice is embedded in the estimation, the independence assumption might not be too restrictive. While I assume that establishments take wages as given, the establishment forms expectations on future wages conditional on current state variables (including the current wage). In general, the evolution of the state variables are all conditional on current state variables and choices, lending a lot of empirical realism to the estimation.

Finally, I would caution against the thought that a fully specified model would yield more credible labor adjustment cost estimates. Each degree of realism entails choosing one specification out of many. The multitude of adjustment cost estimates testifies to the fact that these choices are all but innocuous.

5 Estimation

The estimation proceeds in three steps. I provide detailed information on the estimation in the appendix. First, I need to obtain estimates of the state transition probabilities $F(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, d_{t-1})$ and of the choice probabilities $P(d_t|\mathbf{x}_t^o)$. Second, I solve for the unique fixed point of $\mathbf{W}$ which is a function of $F(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, d_{t-1})$, $P(d_t|\mathbf{x}_t^o)$ and $z(\mathbf{x}_t^o, d_t)$ (see appendix). Third, I maximize the likelihood function based on the choice probabilities for the discrete choices, where the choice probabilities are functions of $\mathbf{z}(\mathbf{x}_t^o, d_t)$, $\tilde{\theta}_z = (1/(\sigma_s\sigma) \quad -1/\sigma \quad -\tau^+/\sigma \quad -\tau^-/\sigma)$, and $\mathbf{W}$. Here the parameter vector $\tilde{\theta}_z$ is distinct from θ because of measurement error in sales (see section 7.1) and because the scale parameter σ of u_t^d can be estimated and does not have to be normalized (see appendix A).

In the data, I observe sales, employment, wages paid to every worker, the number of hires and the number of separations. I assume every worker at any point of time is paid the same wage within an establishment, which I take to be the median wage in the establishment rather than the mean due to some right-censoring issues of the wage data. I discretize the state space into an array of $\omega_w \times \omega_S \times \omega_\ell$ points, where I choose $\omega_w = 15$ for the wage, $\omega_S = 20$ for sales, and $\omega_\ell = 40$ for employment, creating thus an array of 12,000 points. I choose the cell boundaries for w and S to have an equal number of observations in each cell. For ℓ, I choose a fine discretization for small ℓ, and wider intervals for higher levels of employment. For each of these points and each discrete choice, I calculate $P(d_t|\mathbf{x}_t^o)$ by a nearest-neighbor estimator. In particular, I first normalize the state vector (w,S,ℓ) to a mean of zero and an identity variance matrix. Call the normalized data $(\bar{w},\bar{S},\bar{\ell})$ and the normalized state point $\bar{\mathbf{x}}_t^o$. For each normalized $\bar{\mathbf{x}}_t^o$, I calculate the Euclidean distance of this point from the normalized data observations, and choose the k observations with the smallest distance. The relative frequencies of the discrete choices among those k observations are used as estimates for P. I follow Pagan and Ullah (1999) in choosing $k = \sqrt{n}$. The transition probability array $F(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, d_{t-1})$ is estimated in the same vein, but only among observations with the discrete choice of interest. To economise on computation time, and in line with my assumption that wages are exogenous to the establishment, I split the estimation of $F(\mathbf{x}_t^o|\mathbf{x}_{t-1}^o, d_{t-1})$ into $F(w_t|\mathbf{x}_{t-1}^o, d_{t-1})$ and $F((S_t,\ell_t)|\mathbf{x}_{t-1}^o, d_{t-1})$.

The components of the vector $\mathbf{z}(\mathbf{x}^o,d)$ are also estimated by a nearest-neighbor algorithm. I estimate $\mathbb{E}(S|\mathbf{x}^o,d)$, $\mathbb{E}(\ell|\mathbf{x}^o,d)$, $\mathbb{E}(h|\mathbf{x}^o,d)$, and $\mathbb{E}(f|\mathbf{x}^o,d)$ as the average sales, average employment, average hires and average separations among the k observations with discrete choice d and the smallest distance to the normalized state point $\bar{\mathbf{x}}^o$.

Having all these objects in place, I start with an initial guess for the vector $\mathbf{W}(\mathbf{x}_{t+1}^o)$, to iterate on equation 11 until convergence is achieved. I calculate the choice probabilities and the likelihood function. The likelihood function is maximized by choice of $\tilde{\theta}$. The costs of hiring and separating from one worker are given by τ^+ and τ^-. The parameter σ is the scale parameter of the type I extreme value distribution, and determines the variance of the distribution as $\frac{\pi^2}{6}\sigma^2$. It gives a measure of how variant the distribution needs to be to account for the hiring and separation heterogeneity of otherwise similar establishments and/or to account for the choice of d of an establishment which based on its choice specific value functions should have chosen a different alternative. The stronger the model is in predicting the observed choices of the sample establishments, the smaller this variance should be. It is thus maybe comparable in spirit to the mean squared error in a linear regression model.

The reader might have noticed a certain inconsistency in parametrically estimating choice probabilities (call them P_2) relying on the non-parametrically estimated initial choice probabilities (P_1). I follow the methodology in Aguirregabiria and Mira (2002) who show that P itself is a fixed point in the sense that iterating on P will lead to a unique solution of the recursive function of P. Thus, I use P_2 based on the estimated parameters $\tilde{\theta}$, to feed them into a renewed computation of $\mathbf{W}$. Then this $\mathbf{W}$ is used to obtain P_3, and the procedure can be repeated any number of times or until convergence in the parameter vector is achieved. In the estimations this was roughly the case after four iterations.

Given that the final likelihood maximization uses many constructed variables based on a state-space discretization and non-parametric estimation, I do not calculate standard errors of the estimates. Depending on the exact specification, obtaining one set of results on a high-performance computer using 20 nodes takes a few hours. Bootstrapping the standard errors with 500 replications would take me a few weeks. Instead of calculating standard errors, I evaluate the performance of my model by out-of-sample predictions.

5.1 Identification

The likelihood function I am maximizing is the likelihood function of a multinomial logit model, and has a unique maximizer, so the model is identified in that sense. In this subsection I am interested in the intuition of what will determine sign and magnitude of the structural coefficients. I estimate the parameters of this model freely. The data can invalidate this model in a variety of ways. In particular, common sense would require $1/(\sigma_s\sigma) > 0$, $-1/\sigma < 0$, $-\tau^+ < 0$, and $-\tau^- < 0$. That is, sales increase the value of a firm, the scale parameter is positive, paying wages decreases the value of the firm, and labor adjustment is indeed costly.[6] What do the data need to be like to yield these signs? The intuition can be captured – and the exposition simplified – if in this section we ignore the continuation value part of the value function and focus on only one variable. The likelihood of choosing the observed choice c out of a set D in a multinomial logit model

[6] To borrow from Tolstoy, there is only one way in which the model can be "correct", but many ways for it to be "wrong".

is

$$
\begin{aligned}
P(d^* = c) =& \frac{\exp(x_c \beta)}{\sum_{a \in D} \exp(x_a \beta)} \\
=& \frac{1}{1 + \sum_{a \neq c} \exp((x_a - x_c)\beta)}
\end{aligned}
$$

If β is positive, the likelihood function will increase each time that $x_a - x_c < 0$. Thus, if the coefficient on sales is to be positive, the sales value among the chosen alternative needs to exceed the sales value among the non-chosen alternatives sufficiently frequently. In general, the chosen option should dominate the non-chosen alternatives with respect to sales. The reverse is required for the coefficient on the wage bill. Ceteris paribus the wage bill paid under the chosen option needs to be lower than the wage bill under the non-chosen options sufficiently often to guarantee a negative sign on the associated parameter.

Now consider the hiring cost parameter. Note that the *estimated* parameter is $-\tau^+/\sigma$. For the choices H, B, F, and P, the variable values for x_a are respectively

$$
\begin{aligned}
\mathbb{E}(h_t | \mathbf{x}_t^o, d_t = H) \\
\mathbb{E}(h_t | \mathbf{x}_t^o, d_t = B) \\
0 \\
0
\end{aligned}
$$

If the chosen alternative is P, then $(x_a - x_c)$ for $a \in \{H, B, F\}$ is

$$
\begin{aligned}
\mathbb{E}(h_t | \mathbf{x}_t^o, d_t = H) - 0 > 0 \\
\mathbb{E}(h_t | \mathbf{x}_t^o, d_t = B) - 0 > 0 \\
0 - 0 = 0
\end{aligned}
$$

Since $(x_a - x_c)$ is greater than 0 in two, and equal to 0 in one case, the likelihood function will be decreasing in the estimated parameter $-\tau^+/\sigma$, or, given $\sigma > 0$, increasing in τ^+. That is, an inactive firm will contribute to a higher hiring cost parameter. If we only observed establishments which do not adjust, then τ^+ would go towards infinity in an estimation. This agrees with our economic intuition. If we never observed establishments which adjust their employment up or down (or both), then we would conclude that it must be prohibitively costly for them to do so. Similar arguments can be made for the remaining discrete choices.

This has clear implications for the problem of aggregation. First consider time aggregation. The shorter the intervals at which firms are assumed to revise their employment, the more likely they are to be inactive. Any hiring or separation during a month will also be a hiring or separation during the year, but a hiring or separation during the year could still mean eleven months of inactivity compared to one month of activity. As for size aggregation, a firm with thousands of employees is much more likely to be active than a small establishment. Treating the large firm as one entity – with one decision maker for hires and separations – and observing this entity constantly in employment adjustments should lead us to infer that hiring and separating cannot be too costly, at least for small adjustments.

I estimate the model at three different adjustment frequencies: monthly, quarterly, and annual.

6 Alternative model

In addition to my benchmark model, I estimate a model which is more similar to the models in the labor adjustment cost literature. I follow closely the exposition in Cooper et al. (2007).[7] Estimating this model serves the purposes of comparing it to the benchmark model (the estimation builds on the same data), and to what extent temporal aggregation is a problem that extends to other classes of models.

The profit function of an establishment is given by

$$\pi_t = A_t \ell_t^{\alpha} - w_t \ell_t - \tau^+ \mathbb{1}(\Delta\ell_t > 0)\Delta\ell_t + \tau^- \mathbb{1}(\Delta\ell_t < 0)\Delta\ell_t$$

where A is a technology parameter, $\mathbb{1}(\cdot)$ is the indicator function, and $\Delta\ell_t = \ell_t - \hat{\ell}_t$ is the establishment-

[7] Their paper is interested in explaining the correlations between hours, employment, vacancies and unemployment. Since I am interested in labor adjustment costs based on a labor demand model, I differ from Cooper et al. (2007) in the choice of moments to match.

initiated labor adjustment. Wages and A are exogenous, but follow a known stochastic process. Specifically,

$$\ln A_t = c_A + \rho_A \ln A_{t-1} + u_{A,t} \tag{6}$$

where $u_{A,t}$ is normally distributed with mean 0 and variance σ_A^2. Furthermore

$$\ln(A_t/w_t) = c_w + \rho_w \ln(A_{t-1}/w_{t-1}) + u_{w,t} \tag{7}$$

where $u_{w,t}$ is normally distributed with mean 0 and variance σ_w^2. I model the process of (A_t/w_t) rather than only w_t because more productive establishments pay higher wages.[8] The dynamic optimization problem of the establishment is

$$V(A_t,(A_t/w_t),\hat{\ell}_t) = \max_{\ell_t} \pi_t + \beta\mathbb{E}(V(A_{t+1},(A_{t+1}/w_{t+1}),\hat{\ell}_{t+1}|A_t,(A_t/w_t),\ell_t)$$

where $\hat{\ell}_{t+1} = \ell_t + \eta_t$, and η_t are net worker-initiated quits and recalls which happen at the end of the period after production has taken place. The establishment thus chooses its employment, and therefore only its net adjustment, but it might enter the succeeding period with a different number of employees due to quits and recalls (thus, there might still be churning). I therefore also include the following process in the estimation:

$$\hat{\ell}_t = c_\ell + \rho_\ell \hat{\ell}_{t-1} + \rho_\Delta \Delta\ell_{t-1} + u_{\ell,t} \tag{8}$$

where $\Delta\ell_t = \ell_t - \hat{\ell}_t$, and $u_{\ell,t}$ is normally distributed with mean 0 and variance σ_ℓ^2.

6.1 Estimation

The main parameters to be estimated are α, τ^+, and τ^-. The targeted moments are:

1. The average establishment size in number of employees. Given adjustment costs, a higher α implies high labor productivity and high labor demand.

[8] I had more success in matching data moments with this process. Modelling the processes of A and w independently could not account for the high correlation between the two variables.

2. The percentage of establishments which are inactive in a period. Given α, this fraction will be governed by the labor adjustment costs τ^+ and τ^-.

3. The percentage of establishments with positive adjustment among those establishments which are active. This moment will balance the adjustment cost parameters. Given the fraction of adjusting establishments, a higher fraction of expanding firms should result in a lower τ^+ estimate relative to τ^-.

The parameters c_ℓ, ρ_ℓ, ρ_Δ and σ_ℓ in equation 8 are estimated outside of the main estimation algorithm, and are based on linear regressions conditional on different establishment sizes. I distinguish 40 size categories (thus estimating a total of 160 parameters), which is also the size of the discrete grid for employment. The decision space is discretized into 27 points, including all adjustments between -12 and 12 employees (which account for more than 99.5 % of all monthly adjustments) and also -41 (the average adjustment if adjustment is less than -12) and +58 (the average adjustment if adjustment is greater than 12).

To estimate the main parameters, I start with an initial vector of parameter values. Based on α, I infer $\ln A_t$ of the establishments in my data (since sales and employment are observed), and estimate the parameters in equations 6 and 7 through linear regression. The values for $\ln A$ and $\ln(A/w)$ are then discretized and probability transition matrices estimated following the method in Tauchen (1986).

Next, the values and optimal choices associated with each point in the discrete space state are calculated through value function iteration. Starting with a uniform distribution of employment ℓ, productivity $\ln A$, and productivity-wages $\ln(A/w)$ across establishments, I use the decision rules and transition probabilities to iterate on this distribution until convergence is achieved. This distribution and decision rules result in the three simulated counterparts of the targeted moments. The Nelder-Mead algorithm is applied to minimize the weighted sum of the difference between data and simulated moments. The weighting matrix is the identity matrix. The targeted and simulated moments for each adjustment frequency are given in table 3.

It can be seen from the table that the percentage of inactive firms is higher for higher adjustment frequencies. The estimator successfully matches the moments for monthly and quarterly adjustment frequency, but has difficulties in matching the moments – especially the third one – at annual adjustment frequencies.

Table 3: Data and simulated moments

	(1)	(2)	(3)
Adjustment frequency	Monthly	Quarterly	Annual
Target moment	*Average establishment size*		
Data	52.4	52.2	54.0
Simulated	52.4	53.6	55.1
Target moment	*Share inactive*		
Data	61.0	40.2	21.1
Simulated	60.9	40.1	21.6
Target moment	*Share hiring if active*		
Data	48.9	50.9	57.4
Simulated	48.6	50.4	53.2

Note: *Average establishment size* in number of workers. *Share inactive* and *share hiring if active* in percentages.

7 Data

The data are linked employer-employee data from Germany (LIAB longitudinal model version 3), covering the years 1993 to 2007. The data have a survey based employer side, and an administrative employee side. The employer survey is conducted annually through in-place interviews. The sampling unit is an *establishment* (in German *Betrieb*), not a firm as a legal entity. Roughly an establishment is a spatial and commercial unit. A firm might thus have many establishments. The population are all establishments with at least one employee subject to social security contributions. The sampling method is stratified random sampling, with larger establishments (in terms of employment) being oversampled. The survey collects information on annual sales, expenditures on intermediate inputs, employment, investment and many other areas. The employee side of the data comes from the administrative records of the German Employment Agency. Every employee subject to social security contributions must be reported by the employing establishment to the Agency for the purpose of computing and collecting social security contributions. As such, the establishment must also report the exact salary paid to the employee. From this administrative data, I know the beginning and the end date of employment spells of any employee employed by any of the establishments surveyed by the establishment panel. More information on the data can be found in Jacobebbinghaus (2008).

The main advantage of this data is the accurate observance of all employment flows. This is what allows me to use gross rather than net flows and what allows me to include the discrete option of simultaneous hiring and separation (in addition to only hiring and only separation) to the choice set. In choosing my estimation

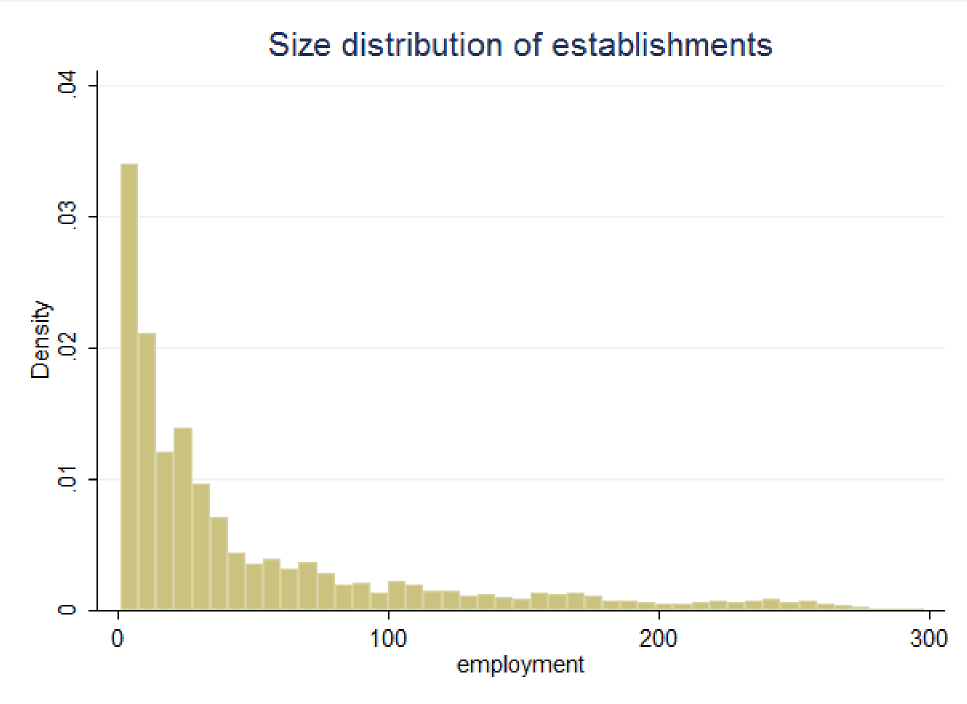

Figure 1: Size distribution

sample I follow the following criteria (the detailed steps of the sample selection are in appendix B): I drop establishments which exhibit large discrepancies in employment between the survey and the administrative records; I choose a balanced panel over four years (1996 - 1999) to abstract from establishment entry and exit, and I include only establishments of 300 or fewer employees to have a more homogeneous sample. Exact definitions of hires, separations and employment are in the appendix.

Finally, I use a randomly selected 90% of the establishment for the estimation, leaving the remaining 10% for out-of-sample predictions. These choices are admittedly somewhat arbitrary. But one has to take a stand regarding how to treat data inconsistencies, and what to count as a – potentially costly – hire and separation. I also estimate the model where hires and separations are defined more loosely and where different criteria are used to classify a separation as worker- or as establishment-initiated. I discuss this in the results section. These steps leave me with 2,816 establishments. For monthly adjustment frequencies I obtain 33,408 establishment-month observations (a few establishments have to be dropped if no employee or no wage is reported for a month). The resulting sample is not representative of all German establishments. But estimating "correct" adjustment costs per se is not the purpose of this paper. Instead I am interested in the effect of different model specifications on adjustment cost estimates.

Figure 1 plots the size (in terms of employment) distribution of the establishments. Most establishments

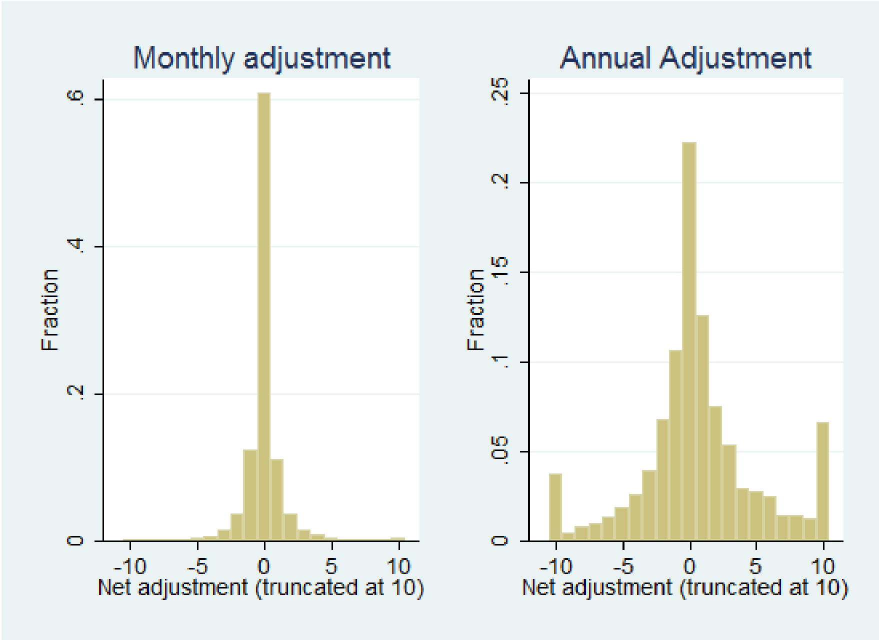

Figure 2: Net adjustment

are small with only a few employees, but the right tail is very long. The more interesting information is shown in figure 2, which shows net labor adjustments for monthly and for annual frequencies. Note that the scales for the two graphs are different. Within a month, about 60% of all establishment-month observations do not change their employment stock (this masks some cases of hires and separations of equal quantity), and almost no establishment has net changes of more than 5 employees. Contrast this with annual adjustments: No adjustment occurs only in little more than 20% of all cases, and the tails are much fatter. There is nothing surprising about this. But the implications for adjustment cost estimates are profound.

Finally, to give a first impression on the importance of distinguishing net from gross adjustments, in table 4 I tabulate the frequencies of each discrete choice for both gross and net changes. We see that even for monthly frequencies net adjustments "hide" many cases (16%) of simultaneous hiring and separations. Importantly, the establishments are not as inactive as one might think observing only net adjustments (55.6% instead of 61%). For annual adjustment, the difference between gross and net adjustment is dramatic. 75.3% of establishments do hire and separate within a year.

7.1 Sales

A substantial problem is posed by the fact that the sales variable is taken from the establishment survey and thus only available as an annual variable. All other relevant variables (hires, separations, wage, and

Table 4: Net vs. gross adjustments (percentage of observations)

	Month		Year	
	Net	Gross	Net	Gross
Only hire (H)	19.2	13.2	44.6	7.7
Only separation (F)	19.8	15.2	33.6	6.8
No hire, no separation (P)	61.0	55.6	21.8	10.2
Both hire and separation (B)	n.a.	15.9	n.a.	75.3
Total	100	100	100	100

employment) can be constructed for any day, but I can not know the revenues for any time interval other than a calendar year. Since I will estimate my model using monthly, quarterly, and annual frequencies, I have to decide how the reported annual sales should be divided unto months and quarters. I follow two approaches. The first is to assume that sales were evenly distributed for the year (*even*). The second is to assume that sales over months and quarters add up to the annual sales number reported by the establishment, that they are never negative, but that changes from period to period are smooth (*smooth*). In particular, taking for example monthly sales, I construct sales to solve the following problem:

$$\begin{aligned}
&\min_{\{s_t\}_{t=1}^{T}} J = \sum_{t=2}^{T}(s_t - s_{t-1})^2 \quad \text{s.t.} \\
&\sum_{t=1}^{12} s_t = S_1 \\
&\sum_{t=13}^{24} s_t = S_2 \\
&\qquad \vdots \\
&\qquad s_t \geq 0 \quad \forall t
\end{aligned}$$

where S are the reported annual sales. Figure 3 depicts the two sales series for an establishment. Of course both series are “wrong”. They contain both classical measurement error due to the establishment probably reporting only a rounded or an approximate sales number, as well as error due to wrongly allocating the annual sales to months and quarters. Moreover, in both of the series there will almost certainly be autocorrelation in the measurement error. But recall that in the estimation the individual sales data are not actually used. Rather, the *average* (over time and establishments) sales conditional on the state and the discrete choice are used, that is $\mathbb{E}(S_t|\mathbf{x}_t^o, d_t)$, so that the errors in the individual sales data should to some extent cancel out. All sales data are net sales, that is revenue minus expenditures on intermediate inputs. This measurement error also explains the reason for the parameter σ_s in equation $\tilde{\theta}$. Remember that I estimate

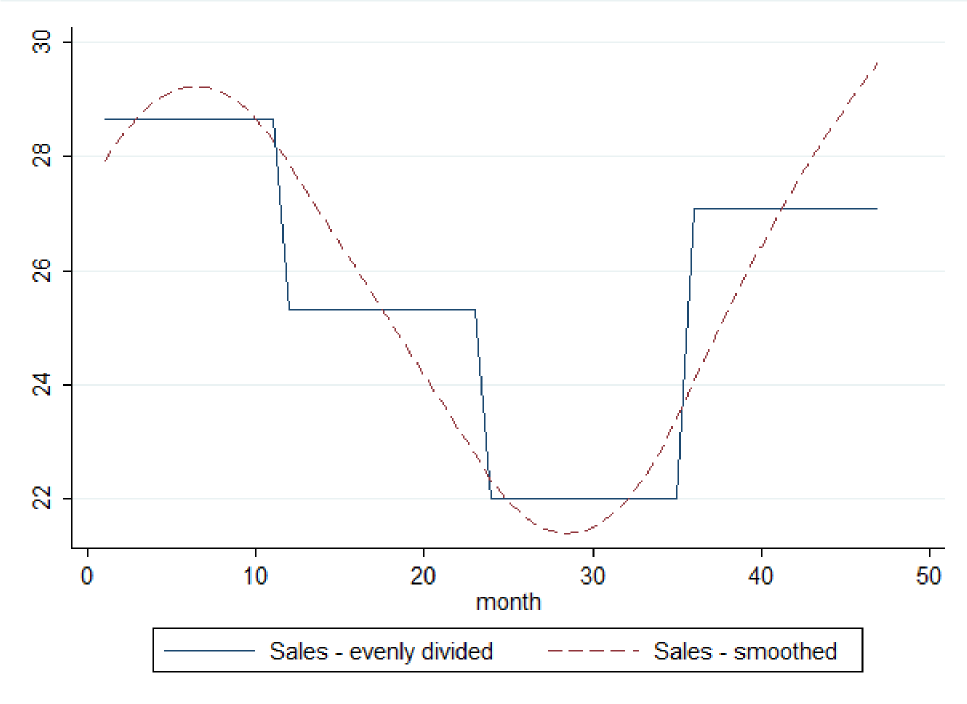

Figure 3: Annual sales divided to months, even and smooth

expected sales $\mathbb{E}(S_t|\mathbf{x}_t^o)$. Suppose the measurement error ζ_t is independent and identically distributed, and that measured sales $\tilde{S}$ relate to true sales S as $S_t = \zeta_t \tilde{S}_t$. Then $\mathbb{E}(S_t|\mathbf{x}_t^o) = \mathbb{E}(\zeta_t \tilde{S}_t|\mathbf{x}_t^o)$. If the covariance between ζ_t and $\tilde{S}_t$ is zero, then we have $\mathbb{E}(S_t|\mathbf{x}_t^o) = \mathbb{E}(\zeta_t \tilde{S}_t|\mathbf{x}_t^o) = \mathbb{E}(\zeta_t)\mathbb{E}(\tilde{S}_t|\mathbf{x}_t^o)$. Thus, $\mathbb{E}(\zeta_t)$ is $1/\sigma_s$ in $\tilde{\theta}$.

7.2 Capital

The literature on factor adjustment costs has mostly assumed that adjustment of all other factors is costless, and the present paper is no exception. Asphjell et al. (2014) and Bloom (2009) are exceptions to this simplification and their results suggest that neglecting capital adjustment costs might seriously bias labor adjustment cost estimates. The reason for why this bias would occur is very intuitive. If capital and labor are interdependent, either as substitutes or complements, adjustments of the two factors are likely to be simultaneous. Thus, costs of investing might easily be attributed to labor adjustment. A model including capital adjustment costs would need to include the capital stock as a state variable. The establishment survey includes the value of annual investments (the same difficulty in allocating this annual figure to months and quarters as we saw for sales would apply). Establishments are also asked what percentage of their investments have been used to replace depreciated capital. With external information on depreciation rates (for example industry-wide) an estimate of the capital stock can be constructed. I constructed capital in this way using depreciation rate information from the German Statistical Office, but the thus constructed variable exhibited within-establishment variations which I judged to be not credible (an AR(1) regression of capital produced an R^2 of 0.05). Furthermore, using this variable would expand the state space from three to four variables, thus diminishing cell sizes for the nearest-neighbor estimates outlined above and multiplying estimation times by at least one order of magnitude. All these considerations have led me to not include capital as a separate variable. However, as a "quick and dirty" alternative I have estimated the model with eight instead of four discrete choices, where each original discrete choice is split into *with investment* and *no investment*, and investment incurs a fixed cost. The choice is *with investment* if in the observed calendar year the establishment reported investment of at least 1,000 euros and *no investment* else. By doing this, I intend to capture some of the contamination that might accrue due to simultaneous capital and labor adjustments. This exercise is just intended as an exploration and I make no claim here that this approach solves the problem of interrelated factor demand.

7.3 Hours

The data distinguish between part- and full-time employment. A more detailed variable on hours worked is not available, and I do not use the part- vs. full-time information. Intuitively, the true adjustment costs will be lower than the ones I estimate. We know that establishments can respond to changing circumstances by changing the average hours worked or by changing the number (and/or composition) of employees. We are not observing the instances in which average hours are adjusted. Presumably, if we were able to close the hours adjustment channel (as in my model), we would see more adjustment through hirings and separations in the data. That is, adjustment through number of employees is less inflexible than it appears in my final sample, because establishments can respond to changing circumstances through both channels, of which I observe only one. Without hours adjustment the establishment would be more active in terms of hiring and separations and estimation would yield lower adjustment costs.

8 Results

The parameter vector to be estimated is $\tilde{\theta}_z = (1/(\sigma_s\sigma) \quad -1/\sigma \quad -\tau^+/\sigma \quad -\tau^-/\sigma)$, where $1/\sigma_s$ is the measurement error in sales, σ is the scale parameter of the extreme value distribution, and τ^+ and τ^- are the

Table 5: Estimated parameters and costs

Discrete choices	Three		Four			Eight
Sales variable	Even	Smooth	Even	Smooth	Smooth	Smooth
	(1)	(2)	(3)	(4)	(5)	(6)
Measurement error sales	0.37	0.34	0.37	0.35	0.20	0.35
Scale parameter	200	204	166	158	227	148
Adjustment Costs in 1,000 Euros						
Hiring cost (τ^+)	73	69	39	34	-48	35
Separation cost (τ^-)	100	98	40	36	-74	36
Fixed cost (τ^f)					445	
Investment cost (τ^i)						1942

Note: *Even* refers to annual sales divided evenly unto the months of the year. *Smooth* refers to annual sales divided unto the months according to the method described in section 7.1. The model with eight discrete choices includes capital adjustment. Each labor adjustment is divided up into with and without capital investment.

labor adjustment costs. It also includes $-\tau^f/\sigma$ if I include a fixed cost of adjustment which is incurred in any period in which hiring or separations take place, or $-\tau^i/\sigma$ if I include a fixed cost of investment as described in the capital section. The main results for monthly adjustment frequencies are presented in table 5. The first two columns are results for models where hiring and separation costs are incurred on net changes. If an establishment had more hirings than separations I considered its choice to be H, in the reverse case F, and if the two variables were equal the choice is P. The choice B – simultaneous hiring and separations – does not exist in this model. By necessity this would be the model when employment is available only as a stock variable, or if labor is treated as homogeneous, in which case the choice B would be economically non-sensible if adjustment is costly. The first column divides annual sales evenly across the months of the year, and the second shows results where sales have been smoothed according to the procedure described in section 7.1.

Qualitatively most results are in line with expectations. The sales measurement error and the scale parameter of the error term are positive, and both hiring and separations are costly. Where the model with net adjustment fails is in delivering adjustment costs in a realistic range. Even though I don't have any priors about these costs,[9] hiring costs of 73,000 Euros per hire and separation costs of 100,000 Euros per separation seem much too high.[10]

[9] The only estimated hiring costs for German establishments are from Mühlemann and Pfeifer (2013) and lie between 4,000 and 6,000 Euros. Harhoff and Kane (1997) report training costs of new apprentices between $5,000 and $10,000.

[10] Both median and mean annual earnings in the data are around 19,000 Euros. The mean is not greater because information on high wages is censored.

Columns three and four add the fourth option B to the model. This decreases the frequency of P choices in the data, since in the 3-choice model some cases in which both hirings and separations occur would have been classified as P. In line with the intuition given in the identification section, fewer inactive periods should decrease the cost parameters, and this is indeed what we observe. The cost parameters in the 4-choice model are 50 to 65 percent smaller than in the 3-choice model. The estimated costs translate to 170 to 180% of the average annual wage of workers. This is a high estimate in comparison to the literature. However, one should keep in mind that the absence of information on hours worked probably overestimates the costs (see the discussion in section 7.3), and that the German labor market is more regulated than the American one.

Hiring and separations seem equally costly, while in the 3-choice model separations are 30% more costly than hirings. This result demonstrates that the distinction between net and gross adjustments is very important for the estimated costs. An interesting result is the relative insensitivity of estimates to which sales measure I use. This is because individual measurement errors due to dividing annual sales largely cancel each other out – irrespective of how the division is carried out – when creating average sales conditional on the state and the discrete choice.

Column 5 adds a fixed – and symmetric – adjustment cost to the model. Unfortunately this model does not work well. The fixed cost of adjustment is extremely high, and every additional hire or separation actually reduces this cost.[11] Finally, the sixth column shows results from a model where the investment choice has been added to the choice set, creating eight discrete choices as described in section 7.2. While investment clearly is an important cost factor, it is surprising that the remaining parameters are only negligibly affected. The adjustment costs are virtually equal to the result of column 4. As mentioned earlier, this last specification is misspecified since capital is not included as a state variable, but the result suggests that investment is orthogonal to the remaining variables of this model.

Table 6 presents results for different assumptions on the frequency of hiring and separation decisions. All dynamic discrete choice models are estimated with four discrete choices, while the method of simulated moments results are based on net adjustment. The first column in the upper panel echoes the result in column 4 of table 5. The second column shows results from the model where decisions are made at the beginning of each quarter, and the third column shows results from annual adjustments. Since sales are reported in annual

[11] All models with fixed adjustment costs have exhibited this result, so I do not pursue this specification any further.

Table 6: Different choice frequencies and estimation methods

Decision frequency	Monthly	Quarterly	Annual	Monthly aggregated
Data frequency	Monthly	Quarterly	Annual	Quarterly
	(1)	(2)	(3)	(4)
Dynamic Discrete Choice				
Scale parameter σ	158	2151	-906	
$-$Hiring cost / σ	-0.21	0.07	0.12	
$-$Separation cost / σ	-0.23	0.11	0.21	
Hiring cost	34	-149	113	
Separation cost	36	-243	190	
Method of Simulated Moments				
Hiring cost	1.5	0.1	-10.9	1.7
Separation cost	-0.5	1.0	16.1	-0.4

Note: All cost estimates in 1,000s of Euros. Sales are smoothed over the periods. The parameter α for the method of simulated moments is estimated to be around 0.5 for all three specifications. Decision frequency refers to how frequently establishment revises its employment, data frequency refers to moments that are being matched, see also table 3. The last column pertains to a monthly model which is aggregated to quarterly data.

frequency, this variable can be taken as it is reported. The final column is only relevant for the alternative model. The model is based on establishments making monthly decisions. These decisions are aggregated to quarterly observations and matched to moments at a quarterly frequency, as inCooper et al. (2007).

In table 6 I have also included the raw estimates of the parameters $-\tau^+/\sigma$ and $-\tau^-/\sigma$. We see that they are increasing in the length of the chosen time period. This is in line with the intuition given in the identification section 5.1. Only the monthly model gives estimates which do not result in the rejection of the model in any dimension. For quarterly data, the finding is one of negative adjustment costs. For annual data, the scale parameter is estimated to be negative. Since this parameter should be strictly positive, the annual adjustment model must be mis-specified in some respect: it might be in assuming that adjustments take place annually, but it might as well be in different respects.

The cost estimates for the alternative model are much smaller and more in line with the estimates obtained by a similar methodology in the literature. The monthly and quarterly models match the target moments closely, while the annual model does not match well the share of hiring establishments among active establishments (table3). What is apparent from the estimates is the sensitivity of the estimates to the second and third data moments (share of inactive establishments and share of hiring establishments among the active ones): Since the share of hiring establishments among the active ones is decreasing in the modelled adjustment frequency (separations are more lumpy than hirings), the costs of hiring relative to separations are also increasing in adjustment frequency. For example, the quarterly model (column 2) has lower hiring and higher separation costs than the monthly model (column 1), so as to accommodate a higher share of hiring firms.

Conditional on adjustment, the estimated cost parameters imply that the cost per adjustment is higher for lower adjustment frequencies (from 0.48 for monthly to 0.60 for annual adjustments). This contradicts the observation of a higher share of inactive establishments at high frequencies. However, the lower share of inactive establishments at low frequencies can easily be explained by the larger marginal profits that would be implied by a longer time horizon. One interesting observation is that a model with monthly adjustment frequency and monthly data can be reproduced if the data is available only quarterly by aggregating up the choices from the monthly model (column 4). The cost estimates of this aggregated model are in the ballpark of the model matched on monthly data. Still, this does not answer the question at which frequency employment decisions are made or best approximated, and the separation cost parameter for the monthly model carries a negative sign. This could be due to sampling error, but the monthly and quarterly estimates lie well below the figures reported in Mühlemann and Pfeifer (2013) and Harhoff and Kane (1997).

The sensitivity of cost estimates to time aggregation is not an entirely new observation (see for examplеBloom (2009)), and not an encouraging finding for the field. It would imply that knowledge of the exact decision making process of managers and establishments must precede any attempt to estimate adjustment costs. This must also be acknowledged as a limitation of this paper: it cannot answer the question which adjustment frequency should be assumed.

Consider next in table 7 how results respond to changes in the model assumptions or sample selection. All results are for monthly adjustment freqencies. Since the two sales variables yield very similar results, I only report results for the smoothed sales variable. I first compare the benchmark model (table 5, column 4) to models with lower discount factors, one with an annual discount factor β of 0.5 (column 2), and one with $\beta = 0$ (column 3), making this a static model, e.g. the firm is not forward looking. Since future sales and wages are discounted, the values associated with labor adjustment are deflated. Naturally, labor adjustment is estimated to be less costly.

Next, I estimate the model defining a separation to be worker-initiated (and therefore costless) if the separation is not followed by another spell within 4 (column 4) or 8 (column 5) weeks, OR if the next ob-

Table 7: Adjustment costs – Alternative specifications (in 1,000 euros)

Specification	(1) Benchmark	(2) $\beta = 0.5$	(3) $\beta = 0.0$	(4) Separations 4 weeks	(5) Separations 8 weeks	(6) Finer employment grid	(7) Small firms	(8) Cost heterogeneity
Hiring cost	34	17	1.2	56	48	33	0	67
Separation cost	36	22	2.9	222	190	36	37	67
Marginal hiring cost								-0.29
Marginal separation cost								-0.24

Note: Estimation frequency is monthly in all cases. Sales are smoothed across months. β in columns 2 and 3 is the value of the time discount factor. In columns 4 and 5 a separation is classified as worker-initiated (and costless) if the separation is not followed by another spell within 4 (8) weeks, OR if the next observed spell is within 4 (8) weeks, and this spell is an employment spell. Column 6 uses 50 instead of 40 discrete points for employment. Column 7 restricts the sample to establishments with no more than 100 employees. In column 8, the hiring and separation costs refer to a wage of 1 Euro, and the marginal costs are changes to the costs for a 1% increase in the wage.

served spell is within 4 or 8 weeks, and this spell is an employment spell. The idea is that an employment spell which soon after its end is followed by another employment spell is a job-to-job transition, and an employment spell followed by neither employment nor unemployment is an out-of-labor-force transition. Thus, only separations which end in unemployment within 4 (8) weeks are considered to be establishment initiated. This results in higher adjustment costs, since this definition of establishment-initiated separations increases the inactive periods of the establishment. Since the inactivity is mostly on the separation dimension, separation cost estimates increase much more than hiring cost estimates.

Column 6 shows results from a model where the length of the employment vector is increased from 40 to 50, that is the employment discretization is finer. It is re-assuring that this hardly affects the estimates. Column 7 shows results from a sample of establishments with at most 100 employees (compared to 300 in the benchmark model). This gives an indication about the effect of aggregation in the firm size. Smaller firms will adjust less frequently, and the cost parameters should be greater in magnitude. This is what we observe in column 7. Finally, I allowed the adjustment costs to be dependent on the log of the median wage paid in the establishment by including the interaction terms $\mathbb{E}(h_t|\mathbf{x}_t^o, d_t \in \{H,B\}) * \log(\bar{w})$ and $\mathbb{E}(f_t|\mathbf{x}_t^o, d_t \in \{F,B\}) * \log(\bar{w})$ into the $\mathbf{z}(\mathbf{x}_t^o, d_t)$ vector. The results in column 8 suggest that if average wages in the establishment are higher, adjustment costs are lower, e.g. a 10% increase in wages decreases hiring costs by 0.29*0.1 = 0.029 thousand (29) Euros, and decreases separation costs by 24 Euros. Establishments which pay higher wages could have lower adjustment costs. Alternatively, establishments which pay higher wages are larger, so that they might be benefiting from economies of scale. Or the estimate for larger establishments understates their true adjustment costs because of aggregation issues discussed in the literature section.

9 Performance

I now turn to an issue which in my opinion in the framework of structural models has remained somewhat unsatisfactory. How can we evaluate/validate a structural model? And against which benchmark? An ideal setting would be to estimate a structural model for a certain time period, and use the estimated model to predict outcomes for another time period in which one of the structural parameters is changed (I will call this "treatment" in line with the treatment effect and evaluation literature). This is the route followed by Aguirregabiria and Alonso-Borrego (2014). That would be informative about the usefulness of the model.

A well-designed, reduced form policy evaluation would give us a credible treatment estimate. Since the unique purpose of the latter is to quantify the treatment effect, the structural modeller could not hope to outperform the reduced form estimate with the predictions of the structural model. Comparing structural models with reduced-form policy evaluations in this way seems misguided, since the purpose of structural models is not primarily, or even if it is, not uniquely, to evaluate a particular treatment. Be that as it may, the present paper does not have a structural change to evaluate.

A sensible evaluation would be based on comparing the goodness of fit (e.g. the Pseudo R-squared based on the likelihood function) and the fraction of correctly predicted outcomes between the structural model and a naive reduced form model. This can be done for the estimation sample, or for out-of-sample observations. What should the naive model be? If I had estimated the structural model based on three outcomes (that is net hiring, net inactivity, and net separations), my model should at least outperform a simple ordinal choice model such as ordinal logit with sales, average wages, and the employment stock (the same state variables as in the structural model) as explanatory variables. But one of the strengths of the present paper is precisely having the fourth choice of both hiring and separations simultaneously, and the four choices do not have a natural ordering. Thus, the best comparison I can make is with respect to a multinomial logit model consisting of the aforementioned four choices, and using the same $\mathbf{z}(\mathbf{x}_t^o, d_t)$ as in the full dynamic model to characterise the choices, but disregarding the dynamic aspect. The test is therefore about whether modelling this problem as a dynamic discrete choice model improves on treating it as static. I present two performance measures: the pseudo R^2 defined as $1 - LL_u/LL_r$, where LL_u is the log-likelihood of the model and LL_r is the log-likelihood of assigning a probability of 25% to each choice; and the fraction of correctly predicted outcomes, where a correct prediction is defined as the observed choice having a higher probability of being chosen than all alternatives. These two measures are calculated for 1) the non-parametric, initial choice probabilities described in the estimation section, 2) the dynamic model, 3) the static model. All measures are calculated for a randomly selected 10% of the original sample which were excluded from the estimation sample.

Table 8 shows the performance measures for these three models. As expected, the non-parametric estimates outperform the parametric predictions. The parametric models are based on four parameters, the non-parametric model separately calculates choice probabilities for 12,000 points in the state space. It serves here as an ideal benchmark. The interesting comparison is between the dynamic and the static model. The

Table 8: Out of sample performance

	(1)	(2)
	Pseudo R^2	Fraction predicted
Non-parametric	0.286	62.6
Dynamic Model	0.070	60.1
Static Model ($\beta = 0$)	0.070	58.9

pseudo R^2 of the dynamic and the static model are equal. The dynamic model does a bit better in predicting the correct discrete choice. The difference in the fraction predicted between the dynamic and static model is significant at 7% if it is tested against the alternative hypothesis that the fraction in the dynamic model is higher. Accounting for the dynamic aspects of an establishment adds little to the model performance in terms of out-of-sample predictions despite having very different estimated adjustment costs – a phenomenon that relates to the poor identification of the discount parameter in these types of models (see Abbring and Daljord (2016) for a treatment of this). The performances are similar because both models fit the data using the same choice and the same state variables. But the estimates are very different because the differences between the values of the different choices are much greater in the dynamic model than in the static model, since the former is composed of the entire expected future stream of profits associated with a choice. This opens up an interesting question of model selection regarding the dynamic aspects of dynamic discrete choice models.

10 Discussion and conclusion

I have analysed labor adjustment costs using a structural dynamic discrete choice model of establishments' hiring and separation decisions. I did this using linked employer-employee data from Germany which allowed me to observe all the in- and out-flows of employees in an establishment, and thus to distinguish net from gross flows. My objective was to analyse which model specifications – in terms of temporal aggregation and net vs. gross adjustments – yield cost estimates in accordance with economic intuition.

Using monthly choice frequencies, the signs of the four parameters are in line with economic intuition and with model assumptions. Hirings and separations are costly, the scale parameter of the extreme value distribution is positive, sales are profitable, and paying wages is unprofitable (ceteris paribus). One of the main objectives of the paper was to show that the timing assumptions, that is the assumed frequency of revising hirings and separations, should matter for the estimated costs, and I have demonstrated that they indeed matter a great deal. The change in the parameters in moving from monthly to less frequent choice frequencies changed the cost parameters in line with my predictions. Indeed, the estimation results clearly rejected quarterly and annual adjustment frequencies.

I also showed that results can vary widely between different estimation strategies. This is clear from the results observed in the literature as well as from a comparison of my benchmark model with the alternative model. However, temporal aggregation is also important in the alternative model.

Another main objective was to highlight the importance between gross and net changes. Again, in line with predictions, I showed that adding the choice of simultaneous hiring and separations result in smaller adjustment costs – the reduction is more than 50%. I also investigated a number of other specifications and sample restrictions and found that adjustment costs are estimated to be higher for smaller establishments, and that a wider definition of worker-initiated separations increases the cost parameters. All these results work through the channel of how much activity is observed. Shorter choice frequencies, classifying fewer separations as establishment initiated, using smaller establishments, and using only net adjustments all result in higher adjustment costs. This is intuitive. Economic conditions change all the time. If firms do not respond frequently by adjusting employment it must be costly to do so. In that sense the present analysis has been very instructive.

However, the dynamic model has barely outperformed a static one but has resulted in very different adjustment cost quantities (e.g. a separation costs 2,000 Euros in the static, and around 35,000 Euros in the dynamic model). Given the similar performances of the static and dynamic model and the very different cost estimates, an open question is how one could discriminate between the two models to reject one in favour of the other.

References

Abbring, Jaap H and Øystein Daljord (2016), "Identifying the discount factor in dynamic discrete choice models." *CEPR Discussion Papers*, 11133.

Abel, Andrew B and Janice C Eberly (1994), "A unified model of investment under uncertainty." *American Economic Review*, 84, 1369–1384.

Abowd, John M, Patrick Corbel, and Francis Kramarz (1999), "The entry and exit of workers and the growth of employment: an analysis of French establishments." *Review of Economics and Statistics*, 81, 170–187.

Abowd, John M and Francis Kramarz (2003), "The costs of hiring and separations." *Labour Economics*, 10, 499–530.

Addison, John T, Pedro Portugal, and José Varejão (2014), "Labor demand research: Toward a better match between better theory and better data." *Labour Economics*, 30, 4–11.

Addison, John T, Claus Schnabel, and Joachim Wagner (2001), "Works councils in Germany: Their effects on establishment performance." *Oxford Economic Papers*, 53, 659–694.

Aguirregabiria, Victor (1997), "Estimation of dynamic programming models with censored dependent variables." *Investigaciones Economicas*, 21, 167–208.

Aguirregabiria, Victor (1999), "The dynamics of markups and inventories in retailing firms." *Review of Economic Studies*, 66, 275–308.

Aguirregabiria, Victor and Cesar Alonso-Borrego (2014), "Labor contracts and flexibility: Evidence from a labor market reform in Spain." *Economic Inquiry*, 52, 930–957.

Aguirregabiria, Victor and Pedro Mira (2002), "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models." *Econometrica*, 70, 1519–1543.

Asphjell, Magne K, Wilko Letterie, Øivind A Nilsen, and Gerard A Pfann (2014), "Sequentiality versus simultaneity: Interrelated factor demand." *Review of Economics and Statistics*, 96, 986–998.

Bachmann, Ronald (2005), "Labour market dynamics in Germany: Hirings, separations, and job-to-job transitions over the business cycle." *SFB 649 Discussion Paper*.

Bachmann, Rüdiger, Christian Bayer, Christian Merkl, Stefan Seth, Heiko Stüber, and Felix Wellschmied (2017), "Worker churn and employment growth at the establishment level." *IZA Discussion Paper*, 11063.

Bentolila, Samuel and Giuseppe Bertola (1990), "Firing costs and labour demand: how bad is eurosclerosis?" *Review of Economic Studies*, 57, 381–402.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 841–890.

Bloom, Nicholas (2009), “The impact of uncertainty shocks.” *Econometrica*, 77, 623–685.

Burgess, Simon, Julia Lane, and David Stevens (2001), “Churning dynamics: an analysis of hires and separations at the employer level.” *Labour Economics*, 8, 1–14.

Cooper, Russell, John Haltiwanger, and Jonathan L Willis (2007), “Search frictions: Matching aggregate and establishment observations.” *Journal of Monetary Economics*, 54, 56–78.

Cooper, Russell, John Haltiwanger, and Jonathan L Willis (2010), “Euler-equation estimation for discrete choice models: A capital accumulation application.” *NBER Discussion Paper*, 15675.

Cooper, Russell, John Haltiwanger, and Jonathan L Willis (2015), “Dynamics of labor demand: Evidence from plant-level observations and aggregate implications.” *Research in Economics*, 69, 37–50.

Ejarque, João Miguel and Pedro Portugal (2007), “Labor adjustment costs in a panel of establishments: A structural approach.” *IZA Discussion Paper*, 3091.

Fallick, Bruce and Charles A Fleischman (2004), “Employer-to-employer flows in the US labor market: The complete picture of gross worker flows.” *mimeo*.

Glitz, Albrecht (2012), “The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in Germany.” *Journal of Labor Economics*, 30, 175–213.

Goolsbee, Austan and David B Gross (2000), “Estimating the form of adjustment costs with data on heterogeneous capital goods.” *mimeo*.

Hall, Robert E (2004), “Measuring factor adjustment costs.” *Quarterly Journal of Economics*, 119, 899–927.

Hamermesh, Daniel S (1989), “Labor demand and the structure of adjustment costs.” *American Economic Review*, 79, 674–689.

Hamermesh, Daniel S and Gerard A Pfann (1996), “Adjustment costs in factor demand.” *Journal of Economic Literature*, 34, 1264–1292.

Harhoff, Dietmar and Thomas J Kane (1997), “Is the German apprenticeship system a panacea for the US labor market?” *Journal of Population Economics*, 10, 171–196.

Hopenhayn, Hugo and Richard Rogerson (1993), "Job turnover and policy evaluation: A general equilibrium analysis." *Journal of Political Economy*, 915–938.

Hotz, V Joseph and Robert A Miller (1993), "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies*, 60, 497–529.

Jacobebbinghaus, Peter (2008), "LIAB Datenhandbuch." *FDZ Datenreport*, 3, 2008.

Jung, Philip and Moritz Kuhn (2014), "Labour market institutions and worker flows: comparing Germany and the US." *The Economic Journal*, 124, 1317–1342.

Jung, Sven and Claus Schnabel (2011), "Paying more than necessary? The wage cushion in Germany." *Labour*, 25, 182–197.

Kohaut, Susanne and Claus Schnabel (2003), "Collective agreements-No thanks!? Extent and determinants of firms' bargaining coverage in Western and Eastern Germany." *Jahrbücher für Nationalökonomie und Statistik*, 223, 312–331.

Kramarz, Francis and Marie-Laure Michaud (2010), "The shape of hiring and separation costs in France." *Labour Economics*, 17, 27–37.

Lapatinas, Athanasios (2009), "Labour adjustment costs: Estimation of a dynamic discrete choice model using panel data for Greek manufacturing firms." *Labour Economics*, 16, 521–533.

Mühlemann, Samuel and Harald Pfeifer (2013), "The structure of hiring costs in Germany: Evidence from firm-level data." *IZA Discussion Paper*, 7656.

Nagypál, Éva (2008), "Worker reallocation over the business cycle: The importance of employer-to-employer transitions." *mimeo*.

Nilsen, Øivind A, Kjell G Salvanes, and Fabio Schiantarelli (2007), "Employment changes, the structure of adjustment costs, and plant size." *European Economic Review*, 51, 577–598.

OECD (1999), *OECD Employment Outlook: June 1999*. Organisation for Economic Co-operation and Development.

Pagan, Adrian and Aman Ullah (1999), *Nonparametric Econometrics*. Cambridge University Press.

Pfann, Gerard A and Franz C Palm (1993), "Asymmetric adjustment costs in non-linear labour demand models for the Netherlands and UK manufacturing sectors." *Review of Economic Studies*, 60, 397–412.

Rota, Paola (2004), "Estimating labor demand with fixed costs." *International Economic Review*, 45, 25–48.

Rust, John (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica*, 55, 999–1033.

Tauchen, George (1986), "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics Letters*, 20, 177–181.

Trapeznikova, Ija (2017), "Employment adjustment and labor utilization." *International Economic Review*, 58, 889–922.

Varejão, José and Pedro Portugal (2007), "Employment dynamics and the structure of labor adjustment costs." *Journal of Labor Economics*, 25, 137–165.

Appendix A - Deriving the likelihood probabilities

The choice-specific value function is given by the current profits of the choice net of adjustment costs and the expected value of next period's value, conditional on the current state and choice.

$$v(\mathbf{x}_t^o, d_t) = \mathbf{z}(\mathbf{x}_t^o, d_t)\theta_z' + u_t^d + \beta\mathbb{E}_{\mathbf{x}_{t+1}^o, \mathbf{u}_{t+1}}\left(V(\mathbf{x}_{t+1}^o, \mathbf{u}_{t+1})|\mathbf{x}_t^o, d_t\right)$$

where $\mathbf{u} = (u^H \quad u^P \quad u^F \quad u^B)$. I can rewrite this as

$$v(\mathbf{x}_t^o, d_t) = \mathbf{z}(\mathbf{x}_t^o, d_t)\theta_z' + u_t^d + \sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o, \mathbf{u}_{t+i}}\left(\mathbf{z}(\mathbf{x}_{t+i}^o, d_{t+i}^*) + u^{d_{t+i}^*}|\mathbf{x}_t^o, d_t\right) \tag{9}$$

I can separately characterize the parts containing $\mathbf{z}$ and u. For $\mathbf{z}$:

$$\begin{aligned}
&\mathbf{z}(\mathbf{x}_t^o, d_t)\theta_z' + \sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o, \mathbf{u}_{t+i}}\left(\mathbf{z}(\mathbf{x}_{t+i}^o, d_{t+i}^*)\theta_z'|\mathbf{x}_t^o, d_t\right) \\
&= \mathbf{z}(\mathbf{x}_t^o, d_t)\theta' + \sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o,|d_t,\mathbf{x}_t^o}\left[\sum_{d_{t+i}\in D}P(d_{t+i}^* = d_{t+i}|\mathbf{x}_{t+i}^o)\mathbf{z}(\mathbf{x}_{t+i}^o, d_{t+i})\theta_z'\right] \\
&= \mathbf{z}(\mathbf{x}_t^o, d_t)\theta_z' + \sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o,|d_t,\mathbf{x}_t^o}\left[\sum_{d_{t+i}\in D}P(d_{t+i}^* = d_{t+i}|\mathbf{x}_{t+i}^o)\mathbf{z}(\mathbf{x}_{t+i}^o, d_{t+i})\theta_z'\right] \\
&= \left(\mathbf{z}(\mathbf{x}_t^o, d_t) + \sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o,|d_t,\mathbf{x}_t^o}\left[\sum_{d_{t+i}\in D}P(d_{t+i}^* = d_{t+i}|\mathbf{x}_{t+i}^o)\mathbf{z}(\mathbf{x}_{t+i}^o, d_{t+i})\right]\right)\theta_z'
\end{aligned}$$

For u:

$$\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o, u_{t+i}}\left(u^{d_{t+i}^*}|\mathbf{x}_t^o, d_t\right)$$

$$\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}^o|\mathbf{x}_t^o, d_t}\left[\sum_{d_{t+i}\in D}P(d_{t+i}^* = d_{t+i}|\mathbf{x}_{t+i}^o)\mathbb{E}\left(u^{d_{t+i}}|\mathbf{x}_{t+i}^o, d_{t+i}^* = d_{t+i}\right)\right] \tag{10}$$

The expression $\mathbb{E}\left(u^{d_{t+i}}|\mathbf{x}_{t+i}^o, d_{t+i}^* = d_{t+i}\right)$ is the expected value of the shock to choice d_{t+i} given that this choice dominated all alternatives. The important insight of Hotz and Miller (1993) is to show that this can be expressed as a function of the choice probabilities. In the case of the type I extreme value distribution, this expectation is given by[12]

$$\mu+\sigma(\gamma-\ln P(d_t|x_t))\equiv \mathbf{e}(\mathbf{x}^o_t,d_t)\theta'_e$$

with $\mathbf{e}(\mathbf{x}^o_t,d_t)=(1 \quad \gamma-\ln P(d_t|\mathbf{x}^o_t))$, $\theta_e=(\mu \quad \sigma)$, and γ is Euler's constant. The parameters of the distribution (μ and σ) are usually not identified and set to 0 and 1 respectively. Replacing this in equation 10 I get

$$\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}^o_{t+i}|\mathbf{x}^o_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}^o_{t+i})\mathbf{e}(\mathbf{x}^o_t,d_t)\theta'_e\right]\times$$
$$\left(\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}^o_{t+i}|\mathbf{x}^o_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}^o_{t+i})\mathbf{e}(\mathbf{x}^o_t,d_t)\right]\right)\theta'_e$$

Define

$$\begin{aligned}\tilde{\mathbf{z}}(\mathbf{x}^o_t,d_t)=&\mathbf{z}(\mathbf{x}^o_t,d_t)+\beta\sum_{\mathbf{x}^o_{t+1}}F(\mathbf{x}^o_{t+1}|\mathbf{x}^o_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}^o_{t+1})\mathbf{z}(\mathbf{x}^o_{t+1},d_{t+1})\right)+\\&\beta^2\sum_{\mathbf{x}^o_{t+1}}F(\mathbf{x}^o_{t+1}|\mathbf{x}^o_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}^o_{t+1})\left[\sum_{\mathbf{x}^o_{t+2}}F(\mathbf{x}^o_{t+2}|\mathbf{x}^o_{t+1},d_{t+1})\left\{\sum_{d_{t+2}\in D}P(d_{t+2}|\mathbf{x}^o_{t+2})\mathbf{z}(\mathbf{x}^o_{t+2},d_{t+2})\right\}\right]\right)\\&+\cdots\end{aligned}$$

Then

$$\begin{aligned}\tilde{\mathbf{z}}(\mathbf{x}^o_t,d_t)=&\mathbf{z}(\mathbf{x}^o_t,d_t)+\beta\sum_{\mathbf{x}^o_{t+1}}F(\mathbf{x}^o_{t+1}|\mathbf{x}^o_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}^o_{t+1})\right.\\&\left.\left[\mathbf{z}(\mathbf{x}^o_{t+1},d_{t+1})+\sum_{k=1}^{\infty}\beta^k\mathbb{E}_{\mathbf{x}^o_{t+1+k}|\mathbf{x}^o_{t+1},d_{t+1}}\left\{\sum_{d_{t+1+k}\in D}P(d_{t+1+k}|\mathbf{x}^o_{t+1+k})\mathbf{z}(\mathbf{x}^o_{t+1+k},d_{t+1+k})\right\}\right]\right)\\=&\mathbf{z}(\mathbf{x}^o_t,d_t)+\beta\sum_{\mathbf{x}^o_{t+1}}F(\mathbf{x}^o_{t+1}|\mathbf{x}^o_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}^o_{t+1})\tilde{\mathbf{z}}(\mathbf{x}^o_{t+1},d_{t+1})\right)\end{aligned}$$

where the last equation follows from

$$\tilde{\mathbf{z}}(\mathbf{x}^o_t,d_t)=\mathbf{z}(\mathbf{x}^o_t,d_t)+\sum_{j=1}^{\infty}\beta^j\mathbb{E}_{\mathbf{x}^o_{t+j}|\mathbf{x}^o_t,d_t}\left(\sum_{d_{t+j}\in D}P(d_{t+j}|\mathbf{x}^o_{t+j})\mathbf{z}(\mathbf{x}^o_{t+j},d_{t+j})\right)$$

[12] I am greatly thankful to Aureo de Paula for providing me with a reference with the proof of this.

Similarly let

$$
\begin{aligned}
\tilde{\mathbf{e}}(\mathbf{x}_t^o, d_t) =&\beta \sum_{\mathbf{x}_{t+1}^o} F(\mathbf{x}_{t+1}^o|\mathbf{x}_t^o, d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1}^o)\mathbf{e}(\mathbf{x}_{t+1}^o, d_{t+1})\right) + \\
&\beta^2 \sum_{\mathbf{x}_{t+1}^o} F(\mathbf{x}_{t+1}^o|\mathbf{x}_t^o, d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1}^o)\left[\sum_{\mathbf{x}_{t+2}^o} F(\mathbf{x}_{t+2}|\mathbf{x}_{t+1}^o, d_{t+1})\left\{\sum_{d_{t+2}\in D} P(d_{t+2}|\mathbf{x}_{t+2}^o)\mathbf{e}(\mathbf{x}_{t+2}^o, d_{t+2})\right\}\right]\right) \\
&+\cdots
\end{aligned}
$$

Then

$$
\begin{aligned}
\tilde{\mathbf{e}}(\mathbf{x}_t^o, d_t) =&\beta \sum_{\mathbf{x}_{t+1}^o} F(\mathbf{x}_{t+1}^o|\mathbf{x}_t^o, d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1}^o)\right. \\
&\left.\left[\mathbf{e}(\mathbf{x}_{t+1}^o, d_{t+1}) + \sum_{k=1}^{\infty} \beta^k \mathbb{E}_{\mathbf{x}_{t+1+k}^o|\mathbf{x}_{t+1}^o, d_{t+1}}\left\{\sum_{d_{t+1+k}\in D} P(d_{t+1+k}|\mathbf{x}_{t+1+k}^o)\mathbf{e}(\mathbf{x}_{t+1+k}^o, d_{t+1+k})\right\}\right]\right) \\
=&\beta \sum_{\mathbf{x}_{t+1}^o} F(\mathbf{x}_{t+1}^o|\mathbf{x}_t^o, d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1}^o)\left\{\mathbf{e}(\mathbf{x}_{t+1}^o, d_{t+1}) + \tilde{\mathbf{e}}(\mathbf{x}_{t+1}^o, d_{t+1})\right\}\right)
\end{aligned}
$$

where the last equation follows from

$$
\tilde{\mathbf{e}}(\mathbf{x}_t^o, d_t) = \sum_{j=1}^{\infty} \beta^j \mathbb{E}_{\mathbf{x}_{t+j}^o|\mathbf{x}_t^o, d_t}\left(\sum_{d_{t+j}\in D} P(d_{t+j}|\mathbf{x}_{t+j}^o)\mathbf{e}(\mathbf{x}_{t+j}^o, d_{t+j})\right)
$$

I rewrite equation 9 as

$$
v(\mathbf{x}_t^o, d_t) = \tilde{\mathbf{z}}(\mathbf{x}_t^o, d_t)\theta_z' + \tilde{\mathbf{e}}(\mathbf{x}_t^o, d_t)\theta_e'
$$

where

Now define

$$
\mathbf{W_z}(\mathbf{x}^o) = \sum_{d\in D} P(d|\mathbf{x}^o)\tilde{\mathbf{z}}(\mathbf{x}^o, d)
$$

$$
\mathbf{W_e}(\mathbf{x}^o) = \sum_{d\in D} P(d|\mathbf{x}^o)\left(\mathbf{e}(\mathbf{x}^o, d) + \tilde{\mathbf{e}}(\mathbf{x}^o, d)\right)
$$

so that

$$\tilde{\mathbf{z}}(\mathbf{x}^o,d) = \mathbf{z}(\mathbf{x}^o,d) + \beta \sum_{\mathbf{x}^{o'}} f(\mathbf{x}^{o'}|\mathbf{x}^o,d)\mathbf{W}_{\mathbf{z}}(\mathbf{x}^{o'})$$

$$\tilde{\mathbf{e}}(\mathbf{x}^o,d) = \beta \sum_{\mathbf{x}^o} f(\mathbf{x}^{o'}|\mathbf{x}^o,d)\mathbf{W}_{\mathbf{e}}(\mathbf{x}^{o'})$$

where $\mathbf{x}^{o'}$ denotes the state vector in the following period. Writing $\mathbf{W}(\mathbf{x}^o) \equiv [\mathbf{W}_{\mathbf{z}}(\mathbf{x}^o)\mathbf{W}_{\mathbf{e}}(\mathbf{x}^o)]$, $\mathbf{W}(\mathbf{x}^o)^*$ is the unique solution to the recursive equation

$$\mathbf{W}(\mathbf{x}^o) = \sum_{d \in D} P(d|\mathbf{x}^o) \times \left([\mathbf{z}(\mathbf{x}^o,d), \mathbf{e}(\mathbf{x}^o,d)] + \beta F(\mathbf{x}^{o'}|\mathbf{x}^o,d) W(\mathbf{x}^{o'}) \right) \tag{11}$$

The choice probability of alternative d is given by

$$P(d|\mathbf{x}^o) = \frac{\exp(v(\mathbf{x}_t^o,d)/\sigma}{\sum_{j=D} \exp(v(\mathbf{x}_t^o,j)/sigma}$$

where σ is the scale parameter of the extreme value distribution. Taking into account the measurement error in sales as described in section 7.1, results in the parameter vectors $\tilde{\theta}_z = (1/(\sigma_s\sigma) \quad -1/\sigma \quad -\tau^+/\sigma \quad -\tau^-/\sigma)$ and $\tilde{\theta}_e = (\mu/\sigma \quad 1)$. Since the parameter μ/σ is multiplied by a constant, it is not identified.

Appendix B - Data cleaning and sample selection

I apply the following sample selection criteria and definitions:

1. The establishment survey reports the stock of employees paying social security for the 30th of June. I compare this reported number to the stock of social security paying employees on the same day from the administrative records. Ideally both numbers should be the same. I drop establishment-year observations where the discrepancy is too large.[13] I also drop establishment-years which report a share of social security paying employees among all employees greater than 100%. Finally I drop establishment-years for which this share is 50% or less *and* $E_{min} \geq 3$.

2. To avoid the additional complication of establishment entry and exit, I use a balanced panel. To have a balanced panel, I face a trade-off between the wide and the longitudinal dimension of the panel. A

[13] Let E_{max} be the greater of the two employment records, and E_{min} the smaller, and let $\Delta \equiv E_{max}/E_{min}$. Any observation with $\Delta \geq 2$ is dropped. In addition, I drop observations if $\Delta \geq 1.66$ *and* $E_{max} \geq 20$, and if $\Delta \geq 1.5$ *and* $E_{max} \geq 50$.

balanced panel for the entire sample period would leave me with too few establishment observations. I pick a panel length of four years, and then select the four consecutive years for which I would have the maximal number of uninterrupted establishment-year observations.

3. I drop establishments which have more than 300 employees in any of the survey years, thus dropping 17% of the remaining establishments. This reduces the representativeness of the sample but also its size heterogeneity (the maximum number of employees in this sample exceeds 20,000). Presumably the excluded establishments (a long right tail) might have very different adjustment costs.

4. Even though I observe the wage paid to every individual employee, the total wage bill at time t is $E(w_t)l_t$. Instead of the mean wage I use the median due to some issues of right-censoring (due to caps to the social security contributions). The mean wage (with censoring) in the sample is 1,620 Euros and the median wage is 1,580 Euros.

5. I define a hire h in period t as an employment spell which starts between the first and last day of the period, *and* if the employee has not been employed by the establishment at any point in the previous 366 days.[14] If the employee has been employed in the establishment in the previous 366 days, I treat him as a recalled employee and implicitly assume that this is done without incurring hiring costs. The recalled worker will still be counted in the employment stock.

6. I define a separation f in period t as an employment spell terminating between the last day of period $t-1$ (e.g. the last day of work is the last day of the finished month) and the day before the last day of t, *and* if the employee is not "recalled" within the next 366 days, *and* if the separation is not from a worker aged 60 or older who does not start a new employment spell subsequently. With this last condition I intend to capture retirements, which I also assume do not cause any separation costs.

7. I define employment at time t to be the number of social security paying employees on the last day of period t.

[14] I also estimate the model with a different set of hiring and separation classifications. See results section.